Performance Modelling and Network Monitoring for Internet of Things (IoT) Connectivity

Sørensen, René Brandborg

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

PERFORMANCE MODELLING AND NETWORK MONITORING FOR INTERNET OF THINGS (IoT) CONNECTIVITY

BY
RENÉ BRANDBORG SØRENSEN

DISSERTATION SUBMITTED 2020

AALBORG UNIVERSITY
DENMARK

Performance Modelling and Network Monitoring for Internet of Things (IoT) Connectivity

Ph.D. Dissertation

René Brandborg Sørensen

Aalborg University
Department of Electronic Systems
Fredrik Bajers Vej 7B
DK-9220 Aalborg

Dissertation submitted:	September 2020
PhD supervisors:	Professor Petar Popovski Aalborg University
	Associate Professor Jimmy Jessen Nielsen Aalborg University
PhD committee:	Associate Professor Rasmus Løvenstein Olsen (chair.) Aalborg University
	Principal Engineer, Director Wael Guibene Charter Communications
	Professor Roberto Verdone University of Bologna
PhD Series:	Technical Faculty of IT and Design, Aalborg University
Department:	Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-817-9

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2020

Abstract

The Internet of Things (IoT) is a term for the next generation of interconnected sensors and actuators, which provide the opportunity for gathering and acting on large sets of data over wide geographical locations. Novel RANs have been developed to serve the various use-cases in IoT, one in particular is the Low Power Wide Area Networks (LPWANs) which aims to connect massive amounts of power-constrained sensors over a wide spatial area. This PhD thesis aims to model the performance of one such novel LPWAN, namely LoRaWAN, and to develop Network Monitoring for IoT networks, that can be used to evaluate the performance of unknown, so called 'black-box', networks without the hassle of modelling the network.

As of 2019, 93 Million LoRa devices were in existence according to Semtech who owns the proprietary LoRa modulation, and the number of devices is predicted to grow immensely in the coming years. This thesis investigates the scalability and Class A end device (ED) performance in terms of latency and outage with specific attention to regulatory requirements for operation in the ISM band and constraints imposed on the performance by the receiver design at the PHY/MAC layer and investigated the potential gain from receiver diversity. The thesis also documents the development of methods for centralised fault-detection in networks with IoT devices that run subroutines, each of which transmits quasi-periodically, such that the aggregate transmissions of all subroutines of IoT device may seem Markovian on first look. These methods detect faults, or outage, based on predictions made after analysis of traffic traces of each device and hierarchical parameterization of quasi-periodic traffic models for each identified subroutine.

Some of the models developed in this thesis are applicable to a wider set of problems than modelling the performance of IoT Networks. More specifically that applies to i) modelling duty-cycling as a queuing problem, ii) approximating the distribution of the maximal number of ongoing interfering transmissions during a transmitted symbol of a reference frame, iii) a framework for modelling G/D/n/n queues and iv) the network monitoring methods, which may be applied to other quasi-periodic processes - for example production line monitoring.

Resumé

Tingenes internet, bedre kendt under det engelske udtryk Internet of Things (IoT), er et term for den næste generation af sammenkoblede sensorer og aktuatorer, der skaber muligheder for at indsamle og agere på baggrund af store datasæt over store geografiske områder. Nye radio netværks teknologier, såkaldte RANs, er blevet udviklet for at servicere de forskellige brugssager, der opstår i tingenes internet. En specifik type netværk er de såkaldte lavenergi bredt netværk, på engelsk low-power wide area networks (LPWANs), der forsøger at tilbyde forbindelse til enorme mængder af batteridrevne enheder over et stort geografisk område. Denne afhandling har til formål at modellere ydelsen af et sådant LPWAN, mere specifikt LoRaWAN, og at udvikle netværk monitorering til IoT netværk, der kan bruges til at evaluere ydeevnen af ukendte netværk, som alternativ til ellers tidskrævende modellering.

Per 2019 eksisterede der 93 millioner LoRa enheder ifølge Semtech, der ejer LoRa modulationen. Antallet af enheder er forudset til at stige stærkt i de kommende år. Denne afhandling undersøger LoRaWANs skalerbarhed og klasse A enheders ydeevne i form af latenstid og afbrydelser med et specifikt fokus på restriktioner fra regulativer og implementeringen af PHY/MAC i modtagerenheden, samt den potentielle gevinst i at anvende flere modtagere. Afhandlingen dokumenterer også udviklingen af metoder til centraliseret fejl-detektion i netværk med IoT enheder, der kører et set ukendte subrutiner, der hver især transmitterer kvasi-peridisk, sådan at den aggregerede trafik for en enkelt enhed kan se tids-invariant ud ved første blik. Disse metoder detekterer fejl og afbrydelser ved at lave forudsigelser for den indkommende trafik, baseret på analyse og hierarkisk parametrisering af trafik modeller for hver sub-rutine på baggrund af den observerede trafik.

Nogle af de modeller, der er udviklet i denne afhandling, er brugbare til et bredere udsnit af problemer end at modellere LoRaWAN. Mere specifikt, så gælder det for i) modellering af duty-cycle som et kø-problem, ii) approksimationen af distributionen for det højeste antal af interfererende transmissioner i et hvilket som helst symbol i en reference transmission, iii) en metodik til modellering af G/D/n/n-køer og iv) netværksovervågningsmetoderne, som kan anvendes til andre kvasi-periodiske processer - for eksempel produktionslinjeovervågning.

Contents

Thesis Details

Thesis Title:	Performance Modelling and Network Monitoring for Internet of Things (IoT) Connectivity
Ph.D. Student:	René Brandborg Sørensen
Supervisors:	Prof. Petar Popovski, Aalborg University
	Assoc. Prof. Jimmy Jessen Nielsen, Aalborg University

The main body of this thesis consist of the following papers.

[A] R. B. Sørensen, D. M. Kim, J. J. Nielsen and P. Popovski, "Analysis of Latency and MAC-Layer Performance for Class A LoRaWAN," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 566–569, Oct. 2017.

[B] R. B. Sørensen, N. Razmi, J. J. Nielsen and P. Popovski, "Analysis of LoRaWAN Uplink with Multiple Demodulating Paths and Capture Effect," *2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2019.

[C] R. B. Sørensen, N. Razmi, J. J. Nielsen and P. Popovski, "On Symbol-wise Collisions and Demodulation Path Blocking in Multi-Gateway LoRaWAN", **submitted, 2020**.

[D] R. B. Sørensen, J. J. Nielsen and P. Popovski, "On multi server queues with degenerate service time distributions and no waiting lines (G/D/n/n),", **submitted, 2020**.

[E] R. B. Sørensen, J. J. Nielsen and P. Popovski, "Machine Learning Methods for Monitoring of Quasi-Periodic Traffic in Massive IoT Networks," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7368–7376, Aug. 2020.

In addition to the main papers, the following publications have also been made.

[1] D. M. Kim, R. B. Sørensen, K. Mahmood, O. N. Osterbo, A. Zanella and P. Popovski, "Data Aggregation and Packet Bundling of Uplink Small Packets for Monitoring Applications in LTE," *IEEE Network*, vol. 31, no. 6, pp. 32–38, 2017.

[2] P. Danzi et al., "Communication Aspects of the Integration of Wireless IoT Devices with Distributed Ledger Technology," *IEEE Network*, vol. 34, no. 1, pp. 47–53, 2020.

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the submitted or published scientific papers which are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty. The thesis is not in its present form acceptable for open publication but only in limited and closed circulation as copyright may not be ensured.

Preface

In truth, I could not have written this thesis without a strong support group. I am truly thankful to already have known the company of so many inspiring people.

I would like to thank to my supervisors Jimmy Jessen Nielsen and Petar Popovski who have been a source of knowledge and inspiration during the PhD, and in Petar's case also throughout my undergraduate and graduate studies. I have to thank Jimmy for his attention to detail and for patiently shielding me against countless poorly formatted subscripts. I would also like to thank my past and present colleagues for the many productive discussions that we have had and without whom lunch would have been a shorter, dull affair.

I would very much also like to extend my appreciation to my dear family, in particular my parents Vivi and Finn whose support and energy know no bounds. Not only did they house me and feed me for much of the PhD, but they have taught me many valuable skills and indeed continue to school me, not least while helping me renovate the home I bought in the early days of the PhD when I still had blue eyes and naive ideas about the workloads of both projects.

Finally, I would like to thank my friends with whom I had the beers over which I shared my joy in successes and the inevitable frustrations I experienced during the PhD.

René Brandborg Sørensen
Aalborg University, September 28, 2020

Part I

Introduction

Introduction

1 Introduction

This subsection presents shortly the motivation for and overall targeted problem that has been investigated and documented throughout the PhD.

1.1 Motivation

The Internet of Things (IoT) is a network of objects consisting of various sensors and actuators, which are interconnected most often through the global internet. The premise of some objects sensing real-world data and others acting upon said data gives rise to a new set of business use-cases and is the foundation for the economy of digital machines. In [] the IoT is rated as having the highest economic impact of several disruptive technologies. The predicted potential ranging between 3.9 Trillion to 11.1 Trillion USD yearly by the year 2025. It is this huge potential and growth that fuels the growth of infrastructure for IoT.

The exchange of information in the IoT is carried out over various networks with wireless Radio Access Networks (RANs) at the front-end in many cases. The traffic characteristics and requirements of machines differ significantly from both conventional voice and broadband traffic, which is mirrored in the two new use-cases specified for 5G: massive Machine Type Communications (mMTC) and Low Latency (LL) & Ultra Reliable (UR) communications. The larger set of machines in the IoT will need mMTC capable networks, which provide connectivity for a large set of nodes, at a relatively low data rate, see [2].

An enabler for mMTC in IoT is Low Power Wide Area Networks (LPWANs) that provide a network, which enables wide area deployments of large numbers of sensors with long battery-life. A variety of LPWANs have been developed in interest of becoming the prevailing technology for IoT, such as SigFox, LoRaWAN and NB-IoT. Close to 100 million LoRa devices existed in 2019 according to [3], while 1.3 billion IoT devices are connected by cellular per [4], most of which are 2G/3G connections and the segment of NB-IoT connected devices is comparable to the number of LoRa devices. The number of

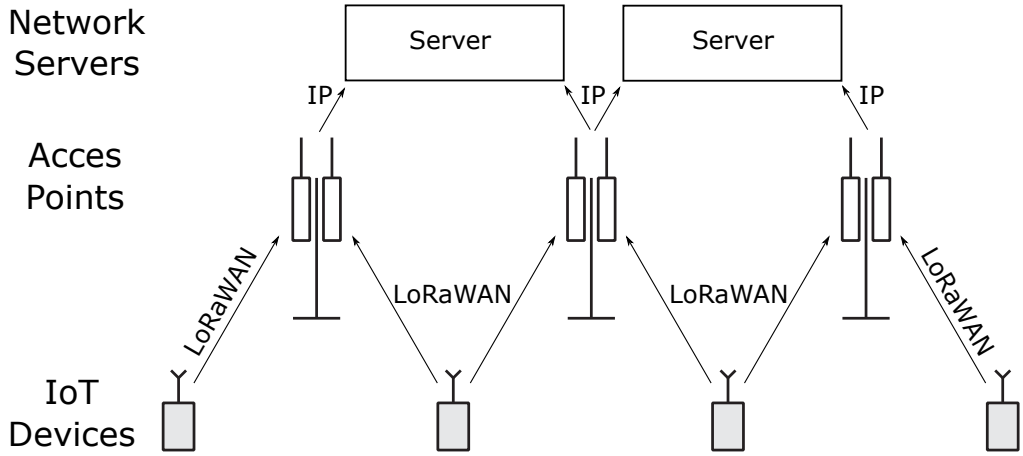


Fig. 1: LoRaWAN network topology. LoRaWAN provides a RAN for EDs to transmit messages to network servers. Several APs in the RAN may receive the same message over LoRaWAN and forward it to the network server through an IP connection. The network server logs the messages from devices and/or forwards them to their end-point.

connected IoT devices is expected to continue to grow, in particular LPWAN connected critical and broadband devices connected by 4G/5G.

The potential of and interest in NB-IoT should not be disregarded, especially given the current predictions of Ericsson, see [4]. Nonetheless, the scope of this thesis is limited to modelling the performance of LoRaWAN, which has also seen vivid growth; Per April 2019 there were 133 commercial LoRaWAN networks in 55 countries. There are 114 certified LoRaWAN devices according to the LoRaWAN Alliance and according to Semtech Corp. the number of LoRa end-nodes was 93 million, see [3], and the number is growing.

1.2 A primer on the LoRaWAN protocol

LoRaWAN is an open standard managed by the LoRa Alliance®. The LoRa Alliance® is a non-profit association of more than 500 member companies. The LoRaWAN standard defines a star-topology RAN that provides long range connectivity at a low bit rate in collaboration with a backbone network comprised of a set of network-servers, see [5]. An access point (AP) serves multiple devices in a star topology, relaying received messages to a network-server. LoRaWAN implements an adaptive data rate (ADR) scheme, which allows the network server to select both the data rate and the channels to be used by each node connected to the AP.

LoRaWAN is based on the LoRa modulation as the PHY layer and LoRa MAC defined in the LoRaWAN specification. The protocol stack for LoRaWAN is depicted in Fig. 2.

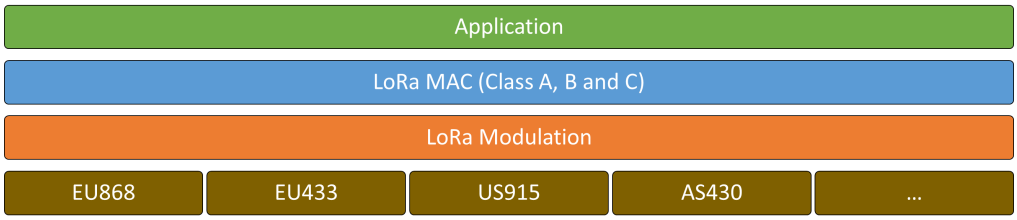


Fig. 2: The protocol stack of LoRaWAN as defined in [5]. The deployment region dictates regulation. Although regulations are depicted at the bottom of the protocol stack here as an abstraction of the wireless medium, coherence with the regional requirements are implemented both in the PHY and MAC layers.

We will clearly distinguish between LoRa as the PHY and the LoRaWAN protocol.

LoRa

LoRa is a proprietary modulation scheme based on Chirp Spread Spectrum (CSS) technique [6]. A chirp is signal transmitted at a constant power, which sweeps over the entire symbol bandwidth (BW) within the symbol period. Symbols are represented by different offsets in the frequency at which the chirp starts and ends. Chirps are insensitive to frequency offsets, bursts of noise and it is a constant envelope modulation, which allows for cheap power-efficient transceivers. The bit rate, chirp rate and symbol rate of LoRa are dictated by the spreading factor (SF), which can be interpreted both as the number of bits transmitted per symbol and a doubling factor of the symbol period. LoRa defines the quasi-orthogonal SFs $m \in M$, $M = \{6, 7, ..., 12\}$. In one symbol period T_S, a chirp covers the entire bandwidth (BW). Thus the symbol rate R_S can be expressed as $R_\mathrm{S} = 1/T_\mathrm{S} = \mathsf{BW}/2^{\mathsf{SF}}$. Taking into account the coding rate (CR), the bit rate R_b, can be expressed as $R_\mathrm{b} = \mathsf{SF} \cdot \mathsf{BW}/2^{\mathsf{SF}} \cdot 4/(4 + \mathsf{CR})$, where CR is selected from 1 through 4.

A LoRa message consists of a preamble, a header, a payload and a CRC for the payload as depicted in Fig. 3. The preamble length can be set to between 6 and 65535 symbols. The preamble also contains a 4.25-symbol length sync word. The sync word consists of 2 upchirp symbols, which are chirps that sweep the BW in a positive manner, and 2.25 downchirp symbols. Thus the total transmission time of the preamble, T_preamble, is given by $T_\mathrm{preamble} = T_\mathrm{s} \cdot (n_\mathrm{preamble} + 4.25)$, where n_preamble is the preamble length. A 20 bit length header is included in explicit header mode to indicate the presence of a CRC for payload and to inform of the payload length and the coding rate. The header is transmitted at the highest coding rate (1/2). In implicit header mode, a header is not transmitted to reduce the time on-air of transmissions, and the transmitter and the receiver must have agreed on a fixed payload length, coding rate and CRC presence beforehand.

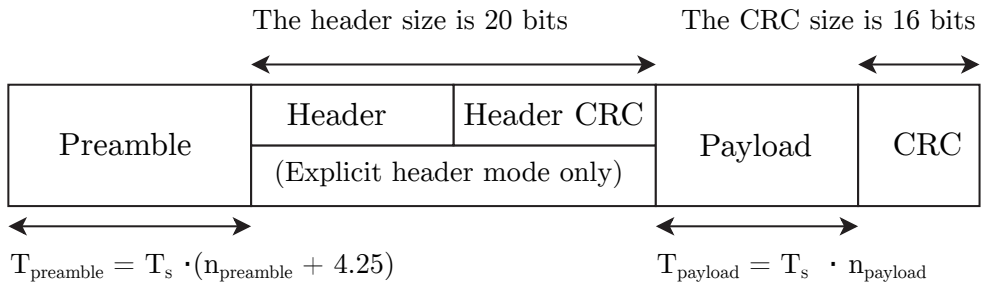


Fig. 3: Physical (PHY) layer LoRa frame structure.

According to Semtech, see [7], the number of symbols needed to transmit a LoRa frame can then be found as

$$n_{\text{packet}_m} = 8 + \max\left(\left\lceil \frac{(8B - 4m + 44)}{(4m - 2 \cdot \mathbb{I}_{\text{DE}})} \right\rceil \cdot (CR + 4), 0\right), \tag{1}$$

where B is the payload of the PHY frame in bytes, $\mathbb{I}_{\text{DE}}$ is a Boolean that indicates whether data rate optimization is used. This feature improves robustness towards clock drift and is mandatory for $m \in \{11, 12\}$.

The total frame length is then $n_{\text{frame}_m} = n_{\text{packet}_m} + n_{\text{preamble}}$.

LoRa MAC

Three different classes of nodes (A, B and C) are defined in LoRaWAN, each of which has a distinct operational mode. Class A has the lowest complexity and energy usage and all LoRaWAN devices are required to implement the class A capability. A class A device operates with grant-less transmissions in the uplink (UL), but can only receive messages in the downlink (DL) within 'receive windows', that are scheduled at a fixed time after a UL transmission finishes. There are two receive windows after a transmission in the uplink. The first window is scheduled to open between 1 to 15 second(s) after the end of an UL transmission with a negligible 20 ms margin of error in the same channel as the UL transmission. The second window opens in a secondary channel 1 second after the end of the first. This operational mode gives class A devices the opportunity to sleep at all times except for a short duration after they transmit in the UL. Class C devices implement the Class A behaviour, but permanently listens for DL transmissions. The receive windows for Class B devices are scheduled by beacons transmitted by a GW whenever a network server queues a DL transmission at the GW. Class B enables polling devices to avoid collisions and initiating DL transmissions from the network server. Class B devices should operate on a separate channel to class A and C devices.

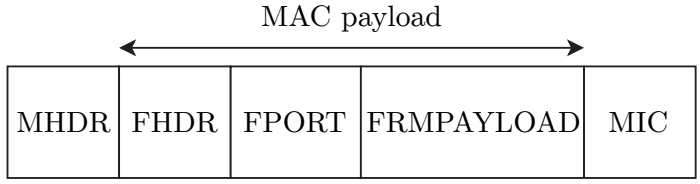


Fig. 4: LoRaWAN (MAC layer) frame structure.

LoRaWAN messages are transmitted as the payload of a LoRa message in Fig. 3. The structure of LoRaWAN messages is shown in Fig. 4. A LoRaWAN message has a MAC header (MHDR field) with 1 byte length specifying message type and version of the encoded frame format. The MAC payload is 1 to B byte(s) where B is dictated by region and spreading factor. The frame header (FHDR) is 6-21 bytes and contains the address of the end-device, a field for frame control, a frame counter and an optional field used to convey MAC commands. A single byte is used for the port field (FPort). The frame payload is 1 to N byte(s) where N is dictated by region and spreading factor. The message integrity code (MIC) in 4 bytes is appended and it is calculated over all fields in the message. In total, the overhead of a LoRaWAN message transmitting a payload with no optional MAC command included is 13 bytes.

Regulatory constraints

LoRaWAN utilizes the industrial, scientific and medical (ISM) radio bands for wireless transmissions, see [8]. ISM bands are unlicensed and subject to regulations in terms of maximum transmit power, duty limiting requirements and bandwidth limitations due to the defined band sizes. In Europe the EU863-870 ISM band uses duty-cycling as a mechanism to regulate usage of the band whereas for example in US and Canada the US902-928 regulations dictate operation in either Frequency-Hopping, Spread-Spectrum (FHSS) or Digital Transmission System (DTS) mode. Notably, India does not restrict the usage of the physical medium beyond a limit on the emitted power. China, Asia, Korea and Russia also implement a duty-cycling mechanism like that implemented in Europe. Europe and Korea also allows for listen before talk (LBT) mechanisms instead of duty-cycling.

The scope of this investigation will be limited to duty-cycling restrictions in particular for the European EU863-870, since duty-cycling is widely allowed around the work, Duty cycling limits the amount of time a device may transmit in the wireless medium by imposing an off-period following transmission that is given by

$$T_{\text{Off}} = T_{\text{On}}\left(\frac{1}{Dc} - 1\right) \tag{2}$$

Table 1: EU863-870 ISM band available sub-bands [9].

Band	f^--f^+ [MHz]	Power [dBm]	Duty cycle	Channels
G	865-868	$\frac{6,2}{100}$[/kHz]	1%	15 (125 kHz) 10 (250 kHz) 4 (500 kHz)
G1	868-868.6	14	1%	3 (125 kHz) 1 (250 kHz)
G2	868.7-869.2	14	0.1%	2 (125 kHz) 1 (250 kHz)
G3	869.4-869.65	27	10%	1 (125 kHz)
G4	869.7-870	7	100%	1 (125 kHz)
G4*	869.7-870	14	1%	1 (125 kHz)

where Dc is the duty cycle. For example, a 1% duty cycle means that if a device will send a frame using a certain sub-band, the same sub-band is prohibited during the next period that corresponds to the on-air duration of that frame multiplied with 99. The device can transmit using the other available sub-band unless entire sub-bands are prevented access. The sub-bands and their power and duty-cycling requirements are listed in Table 1 for the European 863-870 MHz ISM band.

The end-device also obeys another duty cycling mechanism called the aggregated duty cycle, which limits the radio emission of the device. An aggregated duty cycle of 1 corresponds to the device being allowed to transmit at any time, but still in accordance with the regulatory duty cycling. The lowest aggregated duty cycle that can be set, 0, is defined as turning off transmissions entirely for that particular device.

1.3 Thesis Objectives

It is key for both IoT service developers, network operators and network developers to be able to asses the Quality of Service (QoS) of real-life IoT networks and deployments in order to choose the technology that provides the best QoS at a feasible cost.

Prior state of the art

At the beginning of the PhD, the scalability in terms of capacity of LoRaWAN was investigated in [10] where inter-arrival times are implicitly assumed to be fixed. In [11] the scalability is evaluated in terms of goodput and network energy consumption. Duty cycling was pointed out in [12] as a key mechanism in LoRaWAN. Duty cycling is dictated by regulation in many countries as mentioned. It is a central factor limiting the throughput and the latency of a fixed number of devices. The performance limits

of duty-cycled LoRaWAN are pointed out in [13], but only aggregated duty cycle and fixed inter-arrival arrivals are considered.

Performance restrictions in receiver design, capture conditions for LoRa messages, inter-SF interference nor receiver diversity gains were not investigated in [10–13].

However, performance modelling is only one way of investigating the performance; Another is network monitoring, which can provide real-life QoS data based on traffic data from early IoT deployments while also being vital for monitoring and assessing the status of massive numbers of connected IoT devices in order to enable responsible entities to take corrective actions upon failure detection as noted in [14–16].

Research Objectives

This led to the formulation of a set of research questions that was explored through the PhD.

1. Investigate the scalability and QoS of Class A LoRaWAN in terms of the number of supported devices and the latency and outage experienced by devices.
   (a) Identifying and modelling performance trade-offs for LoRaWAN incurred from regulatory requirements of the EU863-870MHz ISM band.
   (b) Identifying and modelling performance trade-offs in LoRaWAN transceiver design at both the PHY and MAC layer.
   (c) Describing the capture conditions for LoRa messages including co- and inter-SF interference and modelling performance trade-offs for different SF-allocations.
   (d) Modelling the performance gain from AP diversity in a LoRaWAN deployment.
2. Develop Network Monitoring applicable to the newly developed and deployed IoT networks intended for a massive number of connections.
   (a) Identify traffic patterns of IoT Networks and shortcomings of conventional Network Monitoring methods for the IoT.
   (b) Develop Network monitoring methods, applicable to the network topology and expected traffic of massive IoT deployments.

1.4 Structure of the Thesis

The thesis is divided into three parts; The section "Thesis contributions" follows after this section to elaborate on the work done throughout the PhD and finally "Conclusion and Further Work" contains reflections and concluding remarks.

2 Thesis contributions

The motivation and main results for each contribution in this thesis are summarized in this chapter. The contributions are divided into two sections. The first section includes the work that aims to model the performance of LoRaWAN deployments and the second section comprises the work done on Network Monitoring for IoT.

We refer to the research questions by the notation (RX.y), for example, (R1.a) refers to the investigation and modelling of regulatory constraints.

Modelling the performance of LoRaWAN networks

2.1 Paper A

Analysis of Latency and MAC-Layer Performance for Class A LoRaWAN, R. B. Sørensen, D. M. Kim, J. J. Nielsen and P. Popovski, *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 566–569, Oct. 2017.

Motivation

The scalability and capacity of Class A LoRaWAN was investigated in [10] and [11] under assumptions of an Aloha network and orthogonality between SFs, in terms of goodput and network energy consumption in the latter and in terms of the number of EDs suppoted in different configurations in the former. The limits of duty-cycled LoRaWAN are pointed out in [13], but only aggregated duty cycle and fixed inter-arrival arrival times are considered.

The regulation in Europe, Asia, China, Russia and Korea dictates duty cycling for individual devices on a per sub-band basis allowing an optional aggregate duty-cycle over all sub-bands bands, see [8]. Semtechs 'LoRaMAC' driver implements duty-cycling over multiple channels and sub-bands so a detailed analysis of the duty-cycling mechanisms impact on traffic division and transmission latency is warranted.

Paper content

This paper introduces a queuing model for the sub-bands and channel candidates for transmission in order to evaluate the internal queuing latency of devices and the aggregate arrival rate from groups of devices in each channel. The model basis is the approximation of a M/D/c-queue based on a M/M/c-queue and it is applicable to both per-sub-band regulatory duty-cycle and the optional aggregated duty-cycle.

Main Results

The main results of this contribution is that we found limits on the throughput and internal queuing latency of individual devices and and the aggregate arrival rate from groups of devices in each channel. We also showed how utilization of multiple sub-bands can increase the throughput and minimizing the latency of individual devices.

This work contributed to the identification of regulatory performance constraints (R1.a) and in particular constraints of the duty-cycling in 'LoRaMAC', Semtech's soft-mac implementation of LoRaWAN (R1.b). Furthermore, the presented model is applicable to any duty-cycled network.

2.2 Paper B

Analysis of LoRaWAN Uplink with Multiple Demodulating Paths and Capture Effect, R. B. Sørensen, N. Razmi, J. J. Nielsen and P. Popovski, *2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2019.

Motivation

Single cell LoRaWAN deployments was investigated using stochastic geometry in [17] under the assumption of orthogonal SFs. The SFs of LoRaWAN were found not to be quasi-orthogonal and capture conditions were given in [18]. Then, [19–21] proceeded to evaluate single cell LoRaWAN using stochastic geometry based on the capture conditions found in [18]. In [19] the allocation of SFs was tuned by assigning SFs to devices based on their distance to the gateway and the reception sensitivities of each SF in order to maximize the average coverage probability.

Another mechanism impacting frame reception was identified experimentally in [22]. Failure was found to happen upon secondary transmission capturing the channel after an earlier transmission had been detected and was being demodulated for the SX1276 LoRa transceiver chipset.

Paper content

We developed a joint spatio-temporal model for LoRaWAN that takes both capture effect and the demodulation capabilities of LoRa transceivers into account by joining the state of art stochastic geometry models for capture effect with Markovian arrival models and queuing theory. The model is based of the SX1301 chipset, which is capable of demodulating 8 frames; however, the model and the methodology can be applied for any architecture. Previous models have not included the temporal-domain constraint given by the number of demodulation paths. Furthermore, we evaluate the performance of single cell LoRaWAN for several SF allocation schemes.

Main Results

We used the developed model to evaluate coverage, throughput and outage in a single GW LoRaWAN deployment and showed that frame dropping due to busy demodulation paths is a non-negligible factor in modelling the performance of LoRaWAN. Additionally, we presented results for several SF-allocation schemes and showed how the SF-allocation scheme is key in determining the throughput, coverage and outage of LoRaWAN deployments.

This work contributed to modelling the demodulation path constraint of the receiver design at the PHY layer (R1.b) and models the performance of LoRaWAN taking into account capture conditions and investigates trade-offs for different SFs (R1.c).

2.3 Paper C

On Symbol-wise Collisions and Demodulation Path Blocking in Multi-Gateway LoRaWAN, R. B. Sørensen, N. Razmi, J. J. Nielsen and P. Popovski, *IEEE Wireless Communications Letters*, **submitted, 2020**.

Motivation

Multiple GW LoRaWAN deployments have been evaluated in [23–25] for various capture and interference conditions. Packet reception in a real-world, multiple GW deployment has been evaluated experimentally in [26]. Interestingly, the maximal RSSI and SNR for a transmission was not found to be among the geographically closest GW-ED pairs, although closer GWs had, on average, better RSSI and received SNR. Also, multiple GWs were found to receive each transmission, sometimes even very remote GWs.

We presented a joint spatio-temporal model in Analysis of LoRaWAN Uplink with Multiple Demodulating Paths and Capture Effect, which takes into account PHY level receiver constraints, which is not taken into account in [23–25]. Furthermore, the temporal interference models of [23–25, 27, 28] evaluate the average number of interferering transmissions during a reference transmission, which is incoherent with actual receiver behavior where symbols are demodulated individually.

Paper content

Here, we took the asynchronism among the interfering frames into account by analyzing the maximal number of ongoing interfering transmissions during any symbol in a reference transmission. This novel model can be used in previous works to evaluate LoRaWAN more realistically under various capture conditions.

In this paper we focused on a simple collision model for a graph-network of EDs connected to GWs with stochastic links. We used this model investigated the benefits in terms of diversity gain from deploying multiple SX1301-based GWs, each operating

under the constraint that the demodulation paths are reserved for decoding ongoing frames.

Main Results

One major result of this work is the good approximation of the distribution of the maximal number of ongoing transmissions during any symbol of a frame. This work modelled the performance gain from AP diversity in a LoRaWAN deployment (R1.d) taking into account capture conditions (R1.c) and modelled the demodulation path constraint of the receiver design at the PHY layer (R1.b). The results showed how diversity gain diminishes both as either more GWs or more EDs are deployed. The model developed in this work is broadly applicable to analysis of any RAN that modulates and demodulates symbols individually.

A quirk in the receiver design was also identified, the preamble detection stage acts as a filter upon the received packet stream, which sorts out colliding or non-capturing packets that arrive with close-inter-arrival times. The resulting packet stream after the detection stage is therefore non-Markovian, so the Erlang B formula does not yield accurate results on blocking for the demodulation stage.

Paper D

On multi server queues with degenerate service time distributions and no waiting lines (G/D/n/n), R. B. Sørensen, J. J. Nielsen and P. Popovski, *IEEE Wireless Communications Letters*, **submitted, 2020**.

Motivation

A distinctive feature of machine operations, is that they are likely to perform different tasks within an almost-deterministic time. This brings relevance to queues with degenerate service times and non-Markovian arrival times. For example, in Paper C we found a LoRa receiver, would filter out colliding transmissions with short inter-arrival times, so that a Markovian arrival process at the front-end of teh receiver would result in a non-Markovian arrival process at the demodulation paths.

Paper content

In this paper, a framework for modelling G/D/n/n queues is presented and we derive a bounds on the blocking probability and close approximation of the blocking probability. Our bounds and approximation are based on an analysis of the counting process for the arrival process rather than cumbersome analysis of the timing of arrivals.

Main Results

Here, we created a model that is applicable in modelling the receiver constraint of LoRa receivers due to having a fixed number of demodulation paths accurately (R1.c).

The main result here is the framework, bounds and approximation itself, which allows for analysis of G/D/n/n queues. We also show how the framework may be applicable to general service times distributions, that is, G/G/n/n queues. The framework is applicable for solving a large set of general problems. Hopefully, the approach taken in our analysis may inspire future research into degenerate queues and general queuing models.

Network Monitoring for IoT

2.4 Paper E

Machine Learning Methods for Monitoring of Quasi-Periodic Traffic in Massive IoT Networks, R. B. Sørensen, J. J. Nielsen and P. Popovski, *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7368–7376, Aug. 2020.

Motivation

Numerous methods for network monitoring in wireless sensor networks (WSNs) exist, see [15, 16] and can be combined in frameworks such as the one in [29], which is based fuzzy logic. The approaches for monitoring can be categorized into 'active', 'passive' and 'hybrid' approaches where active monitoring involves polling devices, passive methods do not add traffic to the network whereas hybrid approaches poll only some devices. Introducing polling traffic for network monitoring in massive IoT deployments would not only create a massive overhead on both UL and DL traffic, but also drain the battery of energy constrained devices, so the potential of passive approaches seems to outweigh the potential of active and hybrid approaches.

In general, the fault detection methods for WSNs assume a PAN multi-hop mesh topology like 6LoWPAN or ZigBee. One example of such a fault detection method is PAD, a passive monitoring method detecting failures based on routing changes that is presented in [30]. Not only is detection of routing changes not applicable in one-hop LPWANs, but PAD and many similar methods add a few bytes of overhead to transmissions, which is non-negligible for energy-constrained devices in low-rate networks. The fault detection methods of [31] and [32] are more interesting since they rely on statistical inference based on the timing of incoming traffic to detect faults in the network. However, both methods also assume that you can gather a set of 'well behaving' data a-priori to starting the fault detection, which only seems feasible in practise for privately owned IoT networks where the traffic models of all devices are well-known a-priori, and in that case faults can be detected in a deterministic manner.

Paper content

We made no assumption about the topology of the network and assumed a quasi-periodic traffic model for IoT devices instead. This approach allowed us to evaluate the state and link performance of individual devices passively by parameterizing the traffic model for each device. This allows for identification of poorly performing and malfunctioning devices in massive IoT deployments so corrective actions can be taken. We developed machine learning methods for parameterizing such traffic models for both devices running 'thin' single application clients and for 'thick' clients running multiple applications. We observed 'thick' client behaviour in a data-set from a LoRaWAN deployment, where

mains-powered sensors and actuators were used to control and manage street lights in rural towns.

We proposed parametric machine learning methods for high resolution, centralised and passive fault detection in arbitrary IoT deployments. The methods use temporal correlations in observed traffic for parameterizing quasi-periodic traffic models. Interestingly, methods from astronomy, where unevenly sampled time series are common, provided inspiration and the methodological basis for the developed algorithms. Specifically, phase-folding [33] inspired the gradient-descent type approach presented called normalised harmonics mean (NHM) while Lomb-Scargle analysis, which is a variant of the classical Fourier periodogram generalised for uneven time-series, see [34–37], was the foundation for a hierarchical clustering approach for parameterizing traffic models for 'thick' clients.

Main Results

This work contributed to the study and development of Network Monitoring for IoT networks (R2.a, R2.b). We developed algorithms practically applicable for fault detection in massive IoT deployments, and investigated the accuracy through simulation and by analysis fo a real-life Smart Metering deployment.

3 Conclusion and Further Work

The main goal of this thesis was to develop tools to ascertain the performance of novel RAN standards intended for usage in massive IoT deployments in particular LoRaWAN, focusing on (i) realistically modelling LoRaWAN by observing transceiver design, regulatory constraints and capture conditions (ii) investigating the scalability of LoRaWAN networks in terms of the supported number of EDs and QoS and (iii) developing network monitoring methods for IoT networks.

We have shown that LoRaWAN scales to an arbitrary number of devices given a large enough set of GWs providing coverage and reception diversity, however at a diminishing return. We found that a single SX1301-based GW can support 1200 Class A EDs spread over 8 channels at an outage of 30% for an average inter-arrival time of 10 minutes of 50 bytes transmissions, given the maximally allowed transmission power and an SF allocation that maximizes the coverage probability throughout the cell. We can scale the number of EDs and the average time between transmissions by any factor to evaluate the number of supported devices at any transmission rate that does not incur non-negligible waiting due to duty-cycling.

We developed network monitoring methods, which can be applied in black-box deployments where neither the network specifications or the parameterized traffic models of devices needs to be known a-priori as long as individual devices are transmitting quasi-periodically. Indeed, these methods allow for ascertaining the performance of individual devices and of the network overall. The latter can be found even if only a subset of devices are transmitting periodically.

We have developed models that do not only model constraints in LoRaWAN, which other works have failed to account for, but they are also applicable to a more general set of problems outside the scope of this thesis. This applies to considering duty-cycling a queuing problem, the approximation of the maximal number of interfering transmissions in any transmitted symbol, the framework for modelling G/D/n/n queues and the network monitoring methods, which may be applied to other quasi-periodic processes - for example production line monitoring.

Future work

There is a fair amount of research questions, which have not been addressed in this thesis. The performance of Class B and Class C LoRaWAN has been modelled in [38, 39], but without considering the restrictions identified in this thesis. Mechanisms such as the adaptive data rate (ADR), the join procedure and broadcasting for eg. firmware over the air (FOTA) are also very interesting areas for study and potentially optimization. The potential of employing new disruptive technologies in GW-receivers in order to to enhance the reception probability is also very interesting, for example SIC as proposed in [40] or MIMO for enhanced coverage.

References

[1] A. Ménard, "How can we recognize the real power of the internet of things?" *McKinsey*, Nov 2017. [Online]. Available: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/how-can-we-recognize-the-real-power-of-the-internet-of-things

[2] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 120–128, 2017.

[3] ABI Research and LoRaWAN Alliance, "Lorawan® and nb-iot: Competitors or complementary?" Tech. Rep.

[4] J. Peter, C. Stephen, B. Greger, K. S. Jason, A. Brian, H. Ahmad, L. Per, and O. Kati, "Ericsson mobility report," Tech. Rep.

[5] LoRa Alliance Technical Committee, *LoRaWAN Specification*, v1.1 ed., LoRa Alliance, Oct. 2017.

[6] "LoRa Modulation Basics," Semtech, Semtech Application Note AN1200.22 Revision 2, May 2015.

[7] Semtech, *SX1276/77/78/79 Datasheet*, rev. 5 ed., Aug. 2016.

[8] LoRa Alliance Technical Committee Regional Parameters Workgroup, *LoRaWAN 1.1 Regional Parameters*, LoRa Alliance, Jan. 2018.

[9] Semtech, "ETSI Compliance of the SX1272/3 LoRa Modem," Semtech, Semtech Application Note AN1200.10 Revision 1, May 2013.

[10] K. Mikhaylov, J. Petäjäjärvi, and T. Haenninen, "Analysis of capacity and scalability of the LoRa low power wide area network technology," in *European Wireless Conference (EW 2016)*, 2016.

[11] M. Bor, U. Roedig, T. Voigt, and J. Alonso, "Do LoRa low-power wide-area networks scale?" in *ACM MSWiM 2016)*, Nov. 2016.

[12] D. Bankov, E. Khorov, and A. Lyakhov, "On the limits of lorawan channel access," in *2016 International Conference on Engineering and Telecommunication (EnT)*, 2016, pp. 10–14.

[13] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, and J. Melia, "Understanding the limits of LoRaWAN," *arXiv preprint arXiv:1607.08011*, 2016.

[14] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, Fourthquarter 2015.

[15] A. Mahapatro and P. M. Khilar, "Fault diagnosis in wireless sensor networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 2000–2026, Fourth 2013.

[16] Z. Zhang, A. Mehmood, L. Shu, Z. Huo, Y. Zhang, and M. Mukherjee, "A survey on fault diagnosis in wireless sensor networks," *IEEE Access*, vol. 6, pp. 11 349–11 364, 2018.

[17] O. Georgiou and U. Raza, "Low power wide area network analysis: Can lora scale?" *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 162–165, April 2017.

[18] D. Croce, M. Gucciardo, S. Mangione, G. Santaromita, and I. Tinnirello, "Impact of lora imperfect orthogonality: Analysis of link-level performance," *IEEE Communications Letters*, vol. 22, no. 4, pp. 796–799, April 2018.

[19] J. Lim and Y. Han, "Spreading factor allocation for massive connectivity in lora systems," *IEEE Communications Letters*, vol. 22, no. 4, pp. 800–803, April 2018.

[20] A. Mahmood, E. G. Sisinni, L. Guntupalli, R. Rondon, S. A. Hassan, and M. Gidlund, "Scalability analysis of a lora network under imperfect orthogonality," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2018.

[21] A. Waret, M. Kaneko, A. Guitton, and N. E. Rachkidy, "Lora throughput analysis with imperfect spreading factor orthogonality," *IEEE Wireless Communications Letters*, pp. 1–1, 2018.

[22] A. Rahmadhani and F. Kuipers, "When lorawan frames collide," *Proc. of the 12th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (ACM WiNTECH 2018)*, 2018.

[23] Z. Qin, Y. Liu, G. Y. Li, and J. A. McCann, "Modelling and analysis of low-power wide-area networks," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.

[24] L. Beltramelli, A. Mahmood, M. Gidlund, P. Österberg, and U. Jennehag, "Interference modelling in a multi-cell lora system," in *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct 2018, pp. 1–8.

[25] M. Ni, M. Jafarizadeh, and R. Zheng, "On the effect of multi-packet reception on redundant gateways in lorawans," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–6.

[26] K. Mikhaylov, M. Stusek, P. Masek, R. Fujdiak, R. Mozny, S. Andreev, and J. Hosek, "On the performance of multi-gateway lorawan deployments: An experimental study," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[27] Z. Li, S. Zozor, J. Drossier, N. Varsier, and Q. Lampin, "2d time-frequency interference modelling using stochastic geometry for performance evaluation in low-power wide-area networks," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.

[28] A. Furtado, J. Pacheco, and R. Oliveira, "Phy/mac uplink performance of lora class a networks," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6528–6538, 2020.

[29] M. Nazari Cheraghlou, A. Khadem-Zadeh, and M. a. Haghparast, "A framework for optimal fault tolerance protocol selection using fuzzy logic oniot sensor layer," *International Journal of Information & Communication Technology Research*, vol. 10, no. 2, 2018. [Online]. Available: http://ijict.itrc.ac.ir/article-1-326-en.html

[30] Y. Liu, K. Liu, and M. Li, "Passive diagnosis for wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1132–1144, Aug 2010.

[31] B. C. Lau, E. W. Ma, and T. W. Chow, "Probabilistic fault detector for wireless sensor network," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3703 – 3711, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417413009548

[32] X. Jin, T. W. S. Chow, Y. Sun, J. Shan, and B. C. P. Lau, "Kuiper test and autoregressive model-based approach for wireless sensor network fault diagnosis," *Wireless Networks*, vol. 21, no. 3, pp. 829–839, Apr 2015. [Online]. Available: https://doi.org/10.1007/s11276-014-0820-0

[33] R. F. Stellingwerf, "Period determination using phase dispersion minimization." *Astrophysical Journal*, vol. 224, pp. 953–960, Sep 1978.

[34] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and Space Science*, vol. 39, no. 2, pp. 447–462, Feb 1976.

[35] J. Scargle, "Studies in astronomical time series analysis. ii - statistical aspects of spectral analysis of unevenly spaced data," *The Astrophysical Journal*, vol. 263, 01 1983.

[36] R. V. Baluev, "Assessing the statistical significance of periodogram peaks," *Monthly Notices of the Royal Astronomical Society*, vol. 385, no. 3, pp. 1279–1285, April 2008.

[37] J. T. VanderPlas, "Understanding the lomb–scargle periodogram," *The Astrophysical Journal Supplement Series*, vol. 236, no. 1, p. 16, may 2018.

[38] P. S. Cheong, J. Bergs, C. Hawinkel, and J. Famaey, "Comparison of lorawan classes and their power consumption," in *2017 IEEE Symposium on Communications and Vehicular Technology (SCVT)*, 2017, pp. 1–6.

[39] F. Delobel, N. El Rachkidy, and A. Guitton, "Analysis of the delay of confirmed downlink frames in class b of lorawan," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, 2017, pp. 1–6.

[40] J. M. d. S. Sant'Ana, A. Hoeller, R. D. Souza, H. Alves, and S. Montejo-Sánchez, "Lora performance analysis with superposed signal decoding," *IEEE Wireless Communications Letters*, pp. 1–1, 2020.

[41] R. B. Sørensen, D. M. Kim, J. J. Nielsen, and P. Popovski, "Analysis of latency and mac-layer performance for class a lorawan," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 566–569, 2017.

[42] R. B. Sørensen, N. Razmi, J. J. Nielsen, and P. Popovski, "Analysis of lorawan uplink with multiple demodulating paths and capture effect," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[43] R. B. Sørensen, J. J. Nielsen, and P. Popovski, "Machine learning methods for monitoring of quasiperiodic traffic in massive iot networks," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7368–7376, 2020.

[44] D. M. Kim, R. B. Sørensen, K. Mahmood, O. N. Osterbo, A. Zanella, and P. Popovski, "Data aggregation and packet bundling of uplink small packets for monitoring applications in lte," *IEEE Network*, vol. 31, no. 6, pp. 32–38, 2017.

[45] P. Danzi, A. E. Kalør, R. B. Sørensen, A. K. Hagelskjaer, L. D. Nguyen, C. Stefanovic, and P. Popovski, "Communication aspects of the integration of wireless iot devices with distributed ledger technology," *IEEE Network*, vol. 34, no. 1, pp. 47–53, 2020.

Part II

Papers

Paper A

Analysis of Latency and MAC-layer Performance for Class A LoRaWAN

René Brandborg Sørensen, Dong Min Kim, Jimmy Jessen Nielsen, Petar Popovski

The paper has been published in the
IEEE Wireless Communications Letters, vol. 6, no. 5, pp. 566–569, 2017

Abstract

We propose analytical models that allow to investigate the performance of Long Range Wide Area Network (LoRaWAN) uplink in terms of latency, collision rate, and throughput under the constraints of the regulatory duty cycling, when assuming exponential inter-arrival times. Our models take into account sub-band selection and the case of sub-band combining. Our numerical evaluations consider specifically the European ISM band, but the analysis is applicable to any coherent band. Protocol simulations are used to validate the proposed models. We find that sub-band selection and combining have a large effect on the QoS experienced in a LoRaWAN cell for a given load. The proposed models allow for optimizing resource allocation within a cell given a set of QoS requirements and a traffic model.

1 Introduction

Services utilizing communications between machines are expected to receive a lot of attention, such as health monitoring, security monitoring and smart grid services [1]. These Internet of Things (IoT) services generate new demands for wireless networks. The spectrum of service scenarios in the IoT is wide and as a result the required quality of service (QoS) across IoT services is also wide. In some scenarios ultra high reliability is required, in others a low latency is required and supporting massive numbers of low-cost and low-complexity devices is still important issue. The devices can be served by the cellular networks and, specifically, by their M2M-evolved versions, such as Narrowband IoT (NB-IoT) [2]. However, there is a low-cost alternative for serving these devices using Low Power Wide Area (LPWA) networks that operate in unlicensed bands. The number of IoT devices connected by non-cellular technologies is expected to grow by 10 billions from 2015 to 2021 [3]. It is therefore of interest to develop QoS models for the LPWA protocols in order to analyze which protocol is best suited for a given service.

Long Range Wide-area Network (LoRaWAN) is an emerging protocol for low-complexity wireless communication in the unlicensed spectrum using Long Range (LoRa) modulation. The scalability and capacity of LoRaWAN is investigated in [4] where it is implicitly assumed that the inter-arrival times are fixed. In [5] the scalability is evaluated in terms of goodput and network energy consumption. One of the key elements of LoRaWAN is the use of duty cycling in order to comply with the requirements for unlicensed operation. Duty cycling is imposed per sub-band by regulation and optionally also aggregated for all bands. It is the central factor that sets limitation on the throughput and the latency of the network. The limits of duty-cycled LoRaWAN are pointed out in [6], but only aggregated duty cycle and fixed inter-arrival arrivals are considered.

The contribution of this paper is an analytical model of the LoRaWAN uplink (UL)

that characterizes the performance, in terms of latency and collision rate, under the influence of regulatory and aggregated duty cycling, assuming exponential inter-arrival times. The obtained latency and collision rate results from the analysis are verified through simulation.

We summarize the key features of LoRaWAN in Section 2. A system model is presented in Section 3 and analysed in Section 4. Numerical results based on the analysis and simulation is shown in Section 5. Concluding remarks are given in Section 6.

2 Long Range Wide Area Network

LoRaWAN is a wireless communication protocol providing long range connectivity at a low bit rate. LoRaWAN is based on the LoRa modulation. LoRaWAN supports LoRa spreading factors 7 to 12. The overhead of a LoRaWAN message with a payload and no optional MAC command included is 13 bytes.

LoRaWAN defines a MAC layer protocol to enable low power wide area networks (LPWAN) [7]. A gateway serves multiple devices in a star topology and relays messages to a central server. LoRaWAN implements an adaptive data rate (ADR) scheme, which allows a network server to select both the data rate and the channels to be used by each node.

Three different classes (A, B and C) of nodes are defined in LoRaWAN. Class A has the lowest complexity and energy usage. All LoRaWAN devices must implement the class A capability. A class A device can receive downlink messages only in a receive window. There are two receive windows after a transmission in the uplink. The first window is scheduled to open 1 to 15 second(s) after the end of an uplink transmission with a negligible 20 ms margin of error. The second window opens 1 second after the end of the first.

LoRaWAN utilizes the industrial, scientific and medical (ISM) radio bands, which are unlicensed and subject to regulations in terms of maximum transmit power, duty cycle and bandwidth. The end-device also obeys a duty cycling mechanism called the aggregated duty cycle, which limits the radio emission of the device. An aggregated duty cycle of 100 % corresponds to the device being allowed to transmit at any time, but still in accordance with the regulatory duty cycling. The lowest aggregated duty cycle of 0 % means that the particular device turns off the transmissions completely.

3 System Model

Consider M devices connected to a single LoRaWAN gateway. Each device is assigned a spreading factor to use for transmission by a network server. We account for the interference through the collision model, where collision occurs when two or more devices try to transmit simultaneously in the same channel using the same spreading factor. We

also consider a LoRa-only configuration, in this work, such that no interference from other technologies is present. Different spreading factors are considered to be entirely orthogonal. A fixed payload size is assumed. We further assume that all devices are class A and have successfully joined the network and transmit the messages without acknowledgement so that there are no downlink transmissions. Due to the absence of acknowledgements, retransmissions are not considered.

Among all sub-bands, a device is given a subset of the sub-bands. Enumerate these sub-bands 1 through c. Let n_i, $i = 1..c$ and δ_i, $i = 1..c$ be the number of channels and the duty-cycle[1] in sub-band i, respectively. As described in the specifications [7] and in the source code of the reference implementation of a LoRa/LoRaWAN device[2], the scheduling of a LoRaWAN transmission happens as follows:

1. A device waits until the end of any receive window.
2. A device waits for any off-period due to aggregated duty cycling.
3. A device checks for available sub-bands, i.e., ones that are not unavailable due to regulatory duty cycling:
   (a) A channel is selected uniformly randomly from the set of channels in all available sub-bands.
   (b) If there is no free sub-band, the transmission is queued in the first free sub-band. A random channel in that sub-band will be selected.

A transmission, limited by the duty cycle δ, with a transmission period T_{tx} infers a holding period, which, including the transmission itself is given by:

$$T_{\text{hold}} = T_{\text{tx}} + T_{\text{tx}} \left(\frac{1}{\delta} - 1 \right) = T_{\text{tx}} \frac{1}{\delta}. \tag{A.1}$$

The *service rate* is the inverse of the holding time, $\mu = \delta / T_{\text{tx}}$. Sub-bands can have different duty cycles and in turn different service rates. Let λ be the generation rate of packets for a device. When several sub-bands are defined for the device the sub-band for the next transmission is selected according to the step 3-a) and 3-b). We define *service ratio* r_i as the fraction of transmissions carried out in the $i-$th sub-band.

4 Analytical Model

In this section the analytical models for latency and collision probability are presented.

[1] δ_i is a normalized value between [0, 1].
[2] https://github.com/Lora-net

4.1 Single Device Model: Latency

The latency of a transmission is the time spent on processing, queueing, transmission of symbols, and propagation. Assuming that the time for processing and propagation are negligible, we hava:eq:

$$T_{\text{total}} = T_{\text{tx}} + T_{\text{w}}. \tag{A.2}$$

We model the wait for reception windows and aggregated duty cycling (steps 1) and 2)) as a single traffic shaping $M/D/1$ queue. The service rate of this $M/D/1$ queue is the slowest mean rate of service in step 1) and 2). For step 3), we model the regulatory duty cycling as an $M/D/c$ queue with heterogeneous servers, where each server corresponds to a sub-band. The waiting time T_{w} for a transmission and the service ratio of each sub-band can then be found from queue theory.

The waiting time, T_w, due to regulatory duty-cycling can be calculated for asymmetric $M/D/c$ queue[3] that models step 3) of the scheduling procedure, but as it is easier to model and compute on a $M/M/c$ queue relative to a $M/D/c$ queue, we use the rule of thumb that the waiting line of a symmetric $M/M/c$ queue is approximately twice that of an $M/D/c$ queue [8] to simplify our analysis. Our simulations show that this is a good approximation also for asymmetric queues.

The waiting time in sub-band i is then:

$$T_{w_i} = \frac{p_{busy,all}}{(\sum_{i=1}^{c} \mu_i + \lambda) \cdot 2}, \tag{A.3}$$

where $p_{busy,all}$ is the Erlang-C probability that all servers are busy. The transmission latency in each band can then be found from Eq. (A.2). The mean latency is given as a weighted sum of the transmission latencies in each sub-band, where the weights are given by the service rate of each sub-band.

The fraction of transmissions in sub-band i, λ_i, is the product of the holding-efficiency of the sub-band (fraction of time it is held) and the service rate of the band throughout that period. Then the service ratio is:

$$r_i = \frac{\mu_i}{\lambda} \cdot (1 - p_{i,\text{idle}}), \tag{A.4}$$

where $p_{i,\text{idle}}$ is the probability that the sub-band i is idle.

The service ratio can be expressed in short-hand forms for the two extreme cases of all sub-bands being available or busy all of the time. When all sub-bands are available at the time of a transmissionthe channel of transmission is selected uniformly from the set of all channels as per step 3-a):

$$\lim_{\lambda \to 0}(r_i) = \frac{n_i}{\sum_{j=1}^{c} n_j}. \tag{A.5}$$

[3]The term “asymmetric" captures the heterogeneous service rates of sub-bands.

In the case that all sub-bands are unavailable at the time of a transmissionthe transmission is carried out in the next available sub-band as per step 3-b):

$$\lim_{\lambda \to \mu_c} (r_i) = \frac{\delta_i}{\sum_{j=1}^{c} \delta_j}. \tag{A.6}$$

In order to describe r_i between these extremes, we must find $p_{i,\text{idle}}$. Hence, we wish to find the steady-state probabilities given a Markov model of the sub-band selection behaviour. For this purpose the model of a *jockeying*[4] $M/M/c$ queue from [9] has been adopted. The Markov model of the jockeying queue has a limited state space since, by definition, the difference in the number of queued transmissions in any two sub-bands may not be larger than one. This allows us to put up a matrix $\mathbf{A}$ containing all state transition probabilities, which can be used to evaluate the steady state probabilities, $\mathbf{P}$, by solving the linear system $\mathbf{A} \cdot \mathbf{P} = 0$. It also allows adoption of a Markov model for LoRaWAN device behaviour, which is step 3) in the sub-band selection, by introducing state transition probabilities based on the number of channels in each non-busy sub-band in $\mathbf{A}$.

The jockeying queue does has a limited state space. As in [9] we approximate the model by making it finite by limiting the queue sizes to 1000. The model now allows us to compute the steady state probabilities of all states; Amongst them $p_{busy,all}$ and $p_{i,\text{idle}}$. Then T_{w_i} and r_i can be calculated from Eq. (A.3) and Eq. (A.4). Note that waiting times are lower for a jockeying queue than a regular queue. Hence applying the rule of thump for approximation of an $M/D/c$ queue from a $M/M/c$ queue on a $M/M/c$: *jockeying* queue, will yield a *lower* latency approximation of the $M/D/c$ queue.

4.2 Multiple Devices Model: Collisions

In this work we assume that no devices are making use of the optional acknowledgement feature of LoRaWAN. Hence there is no DL in the model and as another consequence no retransmissions occur upon collision.

It is empirically found in [10] that spreading factors are not orthogonal in practice and, due to capture effect, one transmission may be received successfully if the power of the wanted transmissions is sufficiently greater than the interfering one. Unfortunately, at present there is no model of capture effect in LoRaWAN and in this work, for simplicity, we assume that all channels and all SFs are orthogonal. When two or more transmissions happen in the same channel, using the same SF, at the same time, they collide. This means we can model the access scheme as multichannel ALOHA random access, as in [4, 6, 11]. Since there are 6 spreading factors defined for LoRaWAN, we have 6 sets

[4] Jockeying: A packet changes queue to a shorter queue if, upon the end of service of another packet, it is located in a longer queue.

of n_i orthogonal Aloha-channels in sub-band i. The collision rate must be evaluated for each spreading factor.

We found the service ratios of each sub-band in Section 4.1. Since the number of devices, the transmission time for the spreading factor being evaluated and the mean inter arrival time are known, we can calculate the load within a sub-band. The load within the sub-band is spread uniformly over the channels allocated to that band. Hence the traffic load of M devices, in sub-band i, given $SF_{i,j}$ is

$$L(i,j) = \frac{\lambda \cdot r_i \cdot T_{\text{tx,j}} \cdot M \cdot p_{\text{SF}_{i,j}}}{n_i} \tag{A.7}$$

where $p_{\text{SF}_{i,j}}$ is the percentage of all devices M, which use the j'th spreading factor in sub-band i.

The collision probability is then

$$p_{\text{col},i,j} = \exp\left(-2 \cdot L(i,j)\right). \tag{A.8}$$

In the paper, only unacknowledged UL transmissions are considered. So DL limitations and retransmissions are not considered in this work. Therefore the outage is caused by collisions can be quantified by our model.

5 Performance Evaluation

In this section the latency given by Eq. (A.2), the service ratios given by Eq. (A.4) and the collision probability given by Eq. (A.8) are evaluated numerically. The evaluation is done for SF 12 based on 125 kHz channels, 50 bytes payload, 13 bytes overhead, code rate 4 and preamble length $n_{\text{preamble}} = 8$.

The latency including the transmission time and the waiting time due to regulatory duty cycling as a function of arrival rate are depicted in Fig. A.1. The latency is plotted for stand-alone usage of each sub-band (G to G4) and for two sub-band combinations (G+G1 and G+G2).

The analytical approximation using Eq. (A.3) for a heterogeneous $M/M/c$ queue provides a tight upper bound of the cases for the multiple sub-bands (G+G1 and G+G2) and a tight approximation for the single band cases. The latency obtained by the jockeying $M/M/c$ queue provides a lower approximation. The results show that lower latencies and higher capacities can be achieved for sub-bands with higher duty-cycles and combinations of bands with high duty-cycles.

The service ratios for the cases with combined sub-bands are plotted in Fig. A.2. We see that combining G with G1 and G2, respectively, leads to very different service ratios for the bands. G contains 15 channels and G1 contains just 3, but they have the same duty-cycle. The combination of G and G1 yields the service ratio limit 15/(15+3) = .834

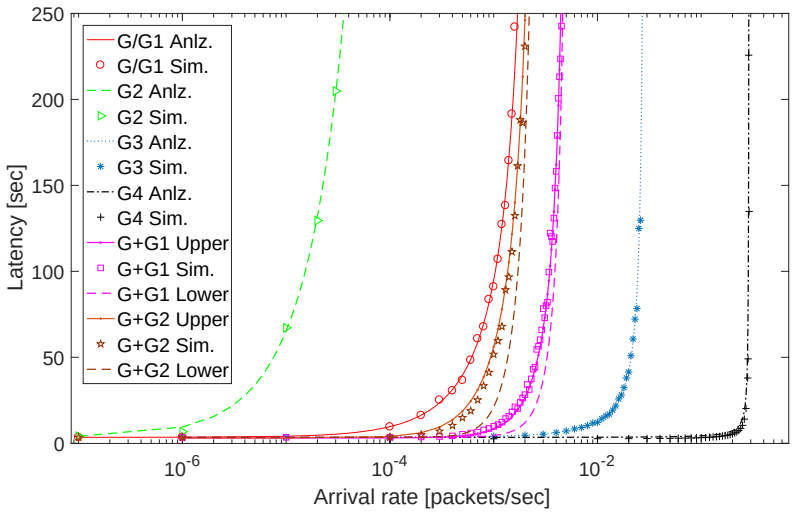


Fig. A.1: ©2017 IEEE: Latencies on all sub-bands and combinations of sub-bands. Results denoted *Upper* and *Lower* are calculated using ordinary $M/M/c$ model and jockeying model, respectively.

for G for low arrival rates, but since the duty-cycling is the same for the sub-bands we have the limit $.01/(.01 + .01)$ for a high arrival rate.

The consequence of the sub-band pairing becomes evident by the collision rates depicted in Fig. A.3. We see that the collision rate for G+G1 is larger than that of G alone or G+G2. This is due to the traffic not being spread equally on the channels for high arrival rates for G+G1. Since the limits of G+G2 are much closer, the load is spread more uniformly over the channels at high arrival rates and we see a drop in collision rate by adding the sub-band. Notice that the devices reach their capacities μ_c before the collision rate comes close to 1. In this way duty-cycling limits the collision rate for each band, allowing for more devices to share the band, but in practice arrivals beyond the capacity of each device would be dropped.

From Fig. A.1 it seems that the sub-band with the highest duty-cycle, G4, is attractive as it delivers low latency even at very high loads. However, when collisions are taken into account, we see that the sub-band has a very high collision rate as it only contains a single sub-channel. On the other hand, the lowest duty-cycle is found in sub-band G2, which has relatively high latency even at low loads, but with a lower collision rate than G4.

In Fig. A.4 the service ratios for G+G4 with an aggregate duty cycle of 0.05 (equivalent to a service rate capacity of the $M/D/1$ queue is 0.0146) and an aggregate duty cycle of 0.075 (equivalent to 0.0219) are plotted. The introduction of the aggregated duty cycle ($M/D/1$ queue) was found to effect the regulatory duty cycle queue ($M/D/c$

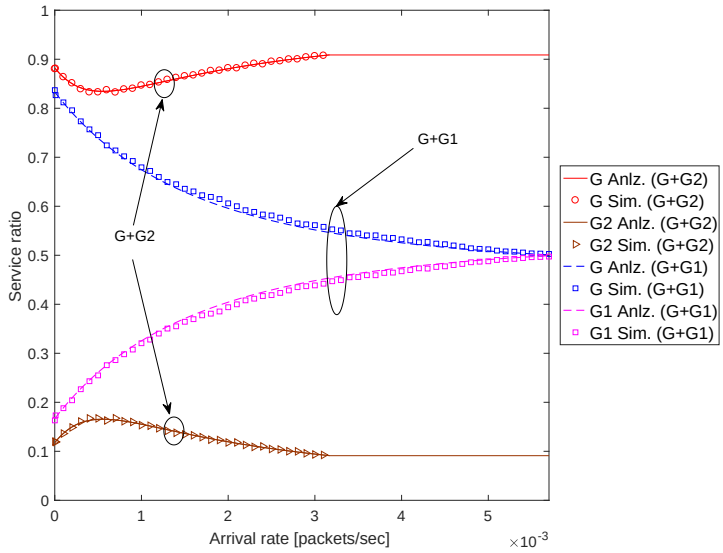


Fig. A.2: ©2017 IEEE: Service ratios for G+G2 and G+G1.

queue) by the service capacity, which freezes the sub-band service ratios of the regulatory queue and limits the obtainable latency.

6 Concluding Remarks

A model for evaluating the performance of LoRaWAN UL in terms of latency and collision probability was presented. The numerical evaluation was done for EU868 ISM band regulations, but the analysis is also valid for other bands utilizing duty cycling, such as the CN779-787 ISM band.

Short-hand forms for the limits of r_i were presented. Equalizing the limits keeps the collision rate of sub-band combining at a minimum. The trade-off for this is a higher latency. The traffic shaping effect of aggregated duty-cycling was shown and may be used as a built-in tool for collision-latency trade-off when combining sub-bands.

The UL model presented in this work, can be combined with DL models for Class A, B and C LoRaWAN devices and more sophisticated collision models to give insight into the bi-directional performance in LoRaWAN.

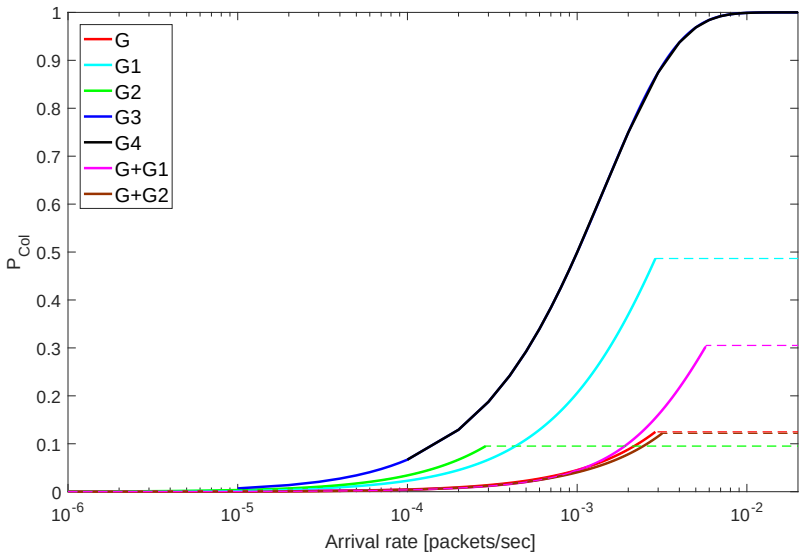


Fig. A.3: ©2017 IEEE: Sub-band collision rates for 100 devices transmitting with SF_{12}.

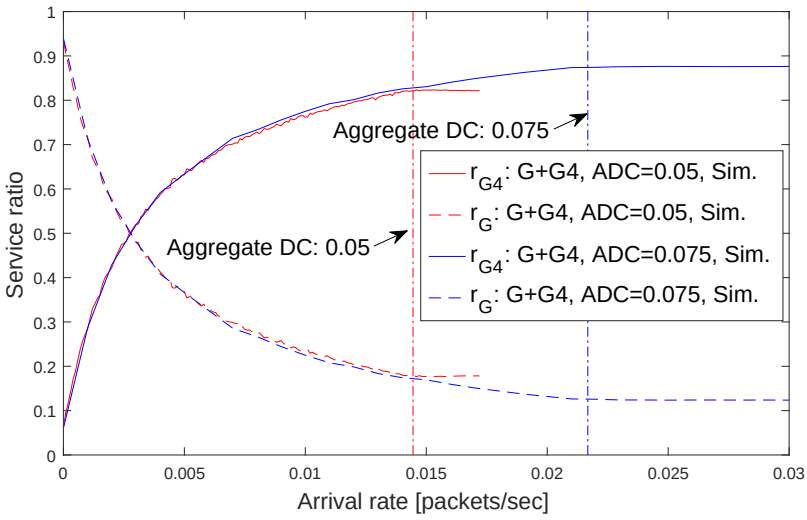


Fig. A.4: ©2017 IEEE: Effect of aggregated duty cycle on service ratios.

References

[1] M. R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel, and L. Ladid, "Internet of things in the 5g era: Enablers, architecture, and business models,"

IEEE Journal on Selected Areas in Communications, vol. 34, no. 3, pp. 510–527, Mar. 2016.

[2] Y.-P. E. Wang, X. Lin, A. Adhikary, A. Grövlen, Y. Sui, Y. Blankenship, J. Bergman, and H. S. Razaghi, "A primer on 3GPP narrowband internet of things (NB-IoT)," *arXiv preprint arXiv:1606.04171*, 2016.

[3] "Ericsson Mobility Report," AB Ericsson, Tech. Rep., June 2016.

[4] K. Mikhaylov, J. Petäjäjärvi, and T. Haenninen, "Analysis of capacity and scalability of the LoRa low power wide area network technology," in *European Wireless Conference (EW 2016)*, 2016.

[5] M. Bor, U. Roedig, T. Voigt, and J. Alonso, "Do LoRa low-power wide-area networks scale?" in *ACM MSWiM 2016)*, Nov. 2016.

[6] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, and J. Melia, "Understanding the limits of LoRaWAN," *arXiv preprint arXiv:1607.08011*, 2016.

[7] N. Sorni, M. Luis, T. Eirich, T. Kramp, and O. Hersent, *LoRaWAN Specification*, v1.0 ed., LoRa Alliance, Jan. 2015.

[8] H. Tijms, "New and old results for the M/D/c queue," *AEU - International Journal of Electronics and Communications*, vol. 60, no. 2, pp. 125–130, 2006.

[9] E. A. Elsayed and A. Bastani, "General solutions of the jockeying problem," *European Journal of Operational Research*, vol. 22, no. 3, pp. 387–396, Dec. 1985.

[10] K. Mikhaylov, J. Petajajarvi, and J. Janhunen, "On LoRaWAN Scalability: Empirical Evaluation of Susceptibility to Inter-Network Interference," *ArXiv e-prints*, Apr. 2017.

[11] A. Augustin, J. Yi, T. Clausen, and W. M. Townsley, "A study of LoRa: Long range & low power networks for the internet of things," *Sensors*, vol. 16, no. 9, 2016.

Paper B

Analysis of LoRaWAN Uplink with Multiple Demodulating Paths and Capture Effect

René Sørensen, Nasrin Razmi, Jimmy Jessen Nielsen, Petar Popovski

The paper has been published in the
Proceedings of IEEE International Conference on Communications (ICC), 2019.

Abstract

Low power wide area networks (LPWANs), such as the ones based on the LoRaWAN protocol, are seen as enablers of large number of IoT applications and services. In this work, we assess the scalability of LoRaWAN by analyzing the frame success probability (FSP) of a LoRa frame while taking into account the capture effect and the number of parallel demodulation paths of the receiving gateway. We have based our model on the commonly used SX1301 gateway chipset, which is capable of demodulating up to eight frames simultaneously; however, the results of the model can be generalized to architectures with arbitrary number of demodulation paths. We have also introduced and investigated three policies for Spreading Factor (SF) allocation. Each policy is evaluated in terms of coverage probability, FSP, and throughput. The overall conclusion is that the presence of multiple demodulation paths introduces a significant change in the analysis and performance of the LoRa random access schemes.

1 Introduction

LoRaWAN is a popular low power wide area network (LPWAN) protocol. It is based on Semtechs' proprietary LoRa modulation, which uses a chirp spread spectrum (CSS) to enable long range and resilient transmissions. The access scheme of LoRaWAN resembles a pure Aloha protocol, however LoRaWAN provides several narrow-band channels and quasi-orthogonal spreading factors (SF) to make LoRaWAN scalable. A considerable research interest has arisen recently about the obtainable link budget of LoRa transmissions, along with studies of capture effect, collisions, and the non-orthogonality of SFs in LoRa-based networks.

There are three mechanisms in LoRaWAN that determine the network performance: coverage, capture effect, and the demodulation capabilities of the receiver. All of these are affected by the SF allocation. The coverage probability is determined by the device location and the sensitivity of the SF. The capture probability is a function of the interference from transmissions using the same SF (co-SF interference) or different SFs (inter-SF interference). Finally the choice of the SF controls the tradeoff between the transmission robustness and collision probability. On one hand, the transmissions that use lower SFs are less robust to noise and interference, resulting in a worsened coverage. On the other hand, lower SF results in a higher rate, which, for fixed payload, translates into a shorter packet duration. This implies that the probability to cause or experience collision (interference) for that packet is lowered for lower SFs. In addition, a transmission with a lower SF occupies a demodulation path in the demodulator for a shorter time interval compared to a transmission with a larger SF.

The initial research in LoRaWAN modeled the protocol as pure Aloha channels for each SF and channel pair as in [1]. This work neglects inter-SF interference and capture

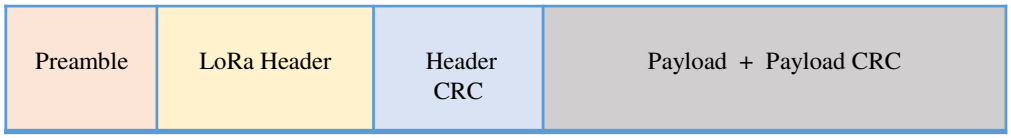


Fig. B.1: ©2019 IEEE: LoRa Frame structure as described in [8]. LoRaWAN frames, consisting of header and payload, are transmitted as the payload of a LoRa frame.

effect. The performance of LoRa has been analyzed and modeled for a single LoRaWAN cell using stochastic geometry and considering capture effect in [2–6]. The authors in [2] model the co-SF interference by considering the interference from the strongest co-SF interferer device. The results show that, the co-SF interference effect increases as the number of devices increases, such that the network becomes interference-limited. The authors in [3] provide a framework to evaluate only the co-SF interference by considering the level of overlap between the interfering packets. Inter-SF interference is modeled with the Co-SF interference in [4–6]. The authors of [4] consider the allocation of SFs in order to maximize the average coverage probability by assigning SFs to devices based on their distance to the gateway and SF sensitivities. The scalability and throughput of LoRaWAN deployments have been evaluated in [5, 6] based on the capture effect and coverage models. Both papers verify that both co-SF and inter-SF transmissions have a considerable effect on the capture probability, and thereby on the LoRaWAN scalability.

A mechanism impacting frame reception beyond coverage and the capture effect was identified experimentally in [7]. The SX1276 LoRa transceiver is found to lock onto frames after detecting four symbols of a preamble. The reception will fail if another frame (even one that captures the channel) begins transmission before the first LoRa frame has been received or certain other timing conditions have been fulfilled.

In this paper, we evaluate the scalability of LoRaWAN analytically by developing a joint model for coverage, capture effect and demodulation capabilities for LoRa transmissions. We build and evaluate our model on the basis of a SX1301 gateway; however, the model and the methodology can be applied for any architecture. The SX1301 is capable of demodulating 8 frames, simultaneously, any transmissions beyond this will be dropped. To our knowledge, this practical limitation has not been included in the previous models from the literature. Finally, we evaluate the performance of several SF allocation schemes based on the developed model.

The rest of paper is organized as follows: In Section 2, we explain the scenario, assumptions, conditions and parameters of our analysis. The analysis is contained within Section 3. Section 4 presents the numerical results and the associated discussion. Finally, the paper is concluded in Section 5.

2 Scenario

We consider a single LoRa channel with a bandwidth $BW = 125$ [kHz]. Each device transmits LoRaWAN frames with payloads of size B bytes at a rate λ_d. A device is assigned a fixed SF m to be used for transmission, where $M = \{7, 8, ..., 12\}$ and $m \in M$ as allowed in LoRaWAN [9]. The symbol period for a given m is $T_{s_m} = \frac{2^m}{BW}$. A LoRA frame uses the LoRaWAN frame as a payload. Each LoRa frame contains a preamble of 10 to 65539 symbols (default is 12), a LoRa header, payload and, optionally, a CRC for the payload as depicted in Fig. B.1. The number of symbols needed for transmission of the LoRa header and the LoRaWAN payload using SF m can be found as (B.1) from [8].

$$n_{\text{pkt}_m} = 8 + \max\left(\left\lceil \frac{(8B - 4m + 44)}{(4m - 2 \cdot \mathbb{I}_{\text{DE}})} \right\rceil \cdot (CR + 4), 0\right), \tag{B.1}$$

where CR is the coding rate. $CR = i$ for $i \in \{1, 2, 3, 4\}$ corresponds to a coding rate equal to $\frac{4}{4+i}$. The indicator $\mathbb{I}_{\text{DE}} = 1$ if low data rate optimization is used and $\mathbb{I}_{\text{DE}} = 0$ otherwise. Low data rate optimization increases robustness towards clock drift and is mandatory for SF $\{11, 12\}$.

The total number of symbols transmitted for a complete frame with SF m is $n_{\text{f}_m} = n_{\text{pre}_m} + n_{\text{pkt}_m}$ where n_{pre_m} denotes the number of symbols in the preamble for m. Then, the total transmission time is $T_{\text{f}_m} = T_{\text{s}_m} \cdot n_{\text{f}_m}$.

2.1 Deployment Model

We consider N devices distributed uniformly in a circular region with radius R with a gateway at the center. The distribution of distance r between the gateway and the devices is defined as $g_R(r)$

$$g_R(r) = \begin{cases} \frac{2r}{R^2} & 0 \leq r \leq R \\ 0 & r \geq R \end{cases} \tag{B.2}$$

The aggregated arrival rate of all devices can be computed as $\lambda = N \cdot \lambda_d$ as long as $\lambda_d < \mu_d$ where μ_d is equal to $\frac{1}{T_{\text{f}_m}/DC}$ and DC is the duty-cycle given as a fraction [10].

Our model is based on SX1301 gateways, which is in widespread use for outdoor LoRaWAN deployments. SX1301 is capable of detecting preambles for all SFs for up to 8 channels at a time. This discrepancy between detection and demodulation is due to the fact that the SX1301 architecture separates the preamble detection from the data acquisition [11, 12].

2.2 Channel model

All devices transmit with power equal to $P_0 = 14$ [dBm]. Let $A(f_c)$ be the deterministic part of the path loss model with carrier frequency f_c and noise power $o_n^2 = -174 + 10\log(BW)$ [dBm], where the noise figure is aassumed to be 0 [dB]. With the assumed bandwidth, we get $A(f_c) = f_c^2 \cdot 10^{-2.8}$. We define c as in [6]

$$c = \frac{P_0 A(f_c)}{o_n^2} \tag{B.3}$$

We denote the required receiver sensitivity for SF m by θ_{RX_m} and let $\gamma_m^{(i)}$, $\gamma_{\mathrm{co},m}^{(i)}$ and $\gamma_{\mathrm{int},m}^{(i)}$ denote the received signal to noise ratio (SNR), the signal to noise plus interference ratios (SINR) of only co-SF interference and only inter-SF interferences for frame i, respectively. $\gamma_{\mathrm{co,int},m}^{(i)}$ denotes the received SINR in the presence of co-SF and inter-SF interferences. $\gamma_m^{(i)}$, $\gamma_{\mathrm{co},m}^{(i)}$, $\gamma_{\mathrm{int},m}^{(i)}$ and $\gamma_{\mathrm{co,int},m}^{(i)}$ are defined as

$$\gamma_m^{(i)} = c|h_i|^2 r_i^{-\alpha}, \tag{B.4}$$

$$\gamma_{\mathrm{co},m}^{(i)} = \frac{|h_i|^2 r_i^{-\alpha}}{\sum_{k \in k_{\mathrm{co},m}} |h_k|^2 r_k^{-\alpha} + \frac{1}{c}}, \tag{B.5}$$

$$\gamma_{\mathrm{int},m}^{(i)} = \frac{|h_i|^2 r_i^{-\alpha}}{\sum_{k \in k_{\mathrm{int},m}} |h_k|^2 r_k^{-\alpha} + \frac{1}{c}}, \tag{B.6}$$

$$\gamma_{\mathrm{co,int},m}^{(i)} = \frac{|h_i|^2 r_i^{-\alpha}}{\sum_{k \in k_{\mathrm{co},m}} |h_k|^2 r_k^{-\alpha} + \sum_{k \in k_{\mathrm{int},m}} |h_k|^2 r_k^{-\alpha} + \frac{1}{c}}, \tag{B.7}$$

where α is the path loss exponent and h is the channel coefficient, which is assumed to be Rayleigh distributed. $k_{\mathrm{co},m}$ and $k_{\mathrm{int},m}$ denote the number of interferer users with the same SF and different SFs, respectively.

2.3 Transmission success

A transmission must capture the channel in order to be successfully received. The channel is captured if the signal passess the thresholds Γ_m, Γ_{co} and $\Gamma_{\mathrm{int},m}$ for capturing the channel with respect to noise, co-SF interference and inter-SF interference, respectively.

The probabilities for these events happening are:

$$\Pr\left(\gamma_m^{(i)} > \Gamma_m\right), \tag{B.8}$$

$$\Pr\left(\gamma_{\text{co,int},m}^{(i)} > \max(\Gamma_{\text{co}}, \Gamma_{\text{int},m})\right), \tag{B.9}$$

$$\Pr\left(\gamma_{\text{co},m}^{(i)} > \Gamma_{\text{co}}\right), \tag{B.10}$$

$$\Pr\left(\gamma_{\text{int},m}^{(i)} > \Gamma_{\text{int},m}\right). \tag{B.11}$$

In the case where there is no inter-SF or co-SF transmission condition 2) reduces to 3) or 4), respectively. Table B.1 lists Γ_m and $\Gamma_{\text{int},m}$ for $m \in M$. Notice that Γ_{co} is 6 [dB], which is always larger than $\Gamma_{\text{int},m}$ so condition 2) simplifies to $\Pr(\gamma_{\text{co,int},m}^{(i)} > \Gamma_{\text{co}})$.

A transmission may be received by the gateway if all the conditions for capturing the channel are fulfilled. However, we must consider the demodulation capability of the receiver which affects the probability of successful reception. If all demodulation paths are busy, then any additional detected frame is dropped. A demodulation path will be assigned to a frame after its preamble has been detected, i.e. four consecutive symbols of the preamble are detected. The header will then be demodulated and if it is correct, so will the rest of the frame. Therefore, the timing of the transmissions must be carefully accounted for in the model.

2.4 SF Allocation

SFs are allocated according to a scheme that is based on annuli, i.e. the radial distance of the device from the BS. Each annuli begins at the radial distance l_{m-1} from the center of the cell and goes to l_m, such that $l_{12} = R$, as depicted in Fig. B.2. Let δ_m denote the fraction of the device population assigned to m and let $g_m(r)$ be the device density distribution for SF m. We define Δ_X as a set of mapping parameters $\Delta_X = \{\{\delta_7, \delta_8, ..., \delta_{12}\}, \{l_7, l_8, ..., l_{12}\}, \{g_7, g_8, ..., g_{12}\}\}$ for the SF allocation scheme X. We present three different allocation schemes and compute Δ in Sec. 3.3.

Table B.1: ©2019 IEEE: Channel capture threshold parameters [6]

SF	θ_{RX_m} [dBm]	Γ_m [dB]	$\Gamma_{\text{int},m}$ [dB]
7	-125	-6	-7.5
8	-128	-9	-9
9	-131	-12	-13.5
10	-134	-15	-15
11	-136	-17.5	-18
12	-137	-20	-22.5

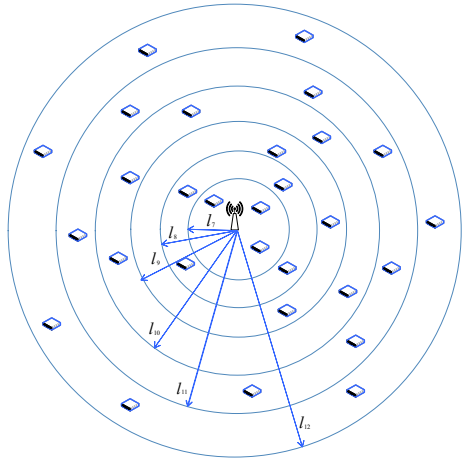


Fig. B.2: ©2019 IEEE: SF allocation scheme based on annuli in a single gateway LoRaWAN cell.

3 Uplink analysis

In this section, we first analyze the FSP by taking into account the capture effect and the timing of collisions. Then, we investigate different SF allocation schemes.

A frame is received successfully if it captures the channel and is not dropped due to all demodulation paths being in use. To evaluate the uplink performance, we derive the probability of a LoRa frame from device i being received successfully, which is denoted by $FSP_m^{(i)}$.

$$FSP_m^{(i)} = FCP_m^{(i)} \cdot (1 - FDP^{(i)}), \tag{B.12}$$

where $FCP_m^{(i)}$ denotes the probability that the frame captures the channel. $FDP^{(i)}$ is the probability a frame being dropped due to all demodulation paths being busy.

To evaluate $FCP_m^{(i)}$ and $FDP^{(i)}$, we need to describe the number of packets which collide with the desired packet. Let $k_{\mathrm{co},m}$ and $k_{\mathrm{int},m}$ denote the number of co-SF and inter-SF frame transmissions, respectively, which interfere with the transmission of frame i using SF m. Notice that we implicitly assume that all interfering frames overlap in time by evaluating the total number of interfering frames over a an entire frame. Realistically, several low SF frames could be placed non-overlapping times within the duration of a high SF frame. Therefore, we derive a lower bound on the $FCP_m^{(i)}$.

We define the traffic load for each source of co-SF and inter-SF interference loads over a period τ by $L_{\mathrm{co},m}(\tau)$ and $L_{\mathrm{int},m}(\tau)$, respectively. When we evaluate the load over a period τ, $L_{\mathrm{co},m}(\tau) = (T_{\mathrm{f}_m} + \tau) \cdot \lambda \cdot \delta_m$ such that we take into account interference from frame transmissions, which had begun before the start of the observed period τ, but

have not ended yet. By equivalent definition, $L_{\text{int},m}(\tau) = \sum_{p \in M, p \neq m}(T_{\text{f}_p} + \tau) \cdot \lambda \cdot \delta_p$ is the inter-SF load where p denotes all SFs different from m. Omitting τ in the notation, the distributions of $k_{\text{co},m}$ and $k_{\text{int},m}$ are given as:

$$P_{k_{\text{co},m}} = \frac{(L_{\text{co},m})^{k_{\text{co},m}} \cdot \exp(-L_{\text{co},m})}{k_{\text{co},m}!}. \tag{B.13}$$

$$P_{k_{\text{int},m}} = \frac{(L_{\text{int},m})^{k_{\text{int},m}} \cdot \exp(-L_{\text{int},m})}{k_{\text{int},m}!}. \tag{B.14}$$

3.1 Derivation of FCP_m^i

To derive the probability of capturing the channel, we evaluate the 4 conditions for capture effect from Sec. 2.3

$k_{\mathbf{co},m} + k_{\mathbf{int},m} = 0$

The probability of a single transmission being transmitted in the channel, is the probability of the devices in our deployment not generating a packet within two packet times.

$$P_{\text{nocol}_m} = \exp\left(-\sum_{p \in M}(T_{\text{f}_m} + T_{\text{f}_p}) \cdot \delta_p \cdot \lambda\right). \tag{B.15}$$

Then the probability of successfully transmitting the frame is equal to

$$P^{(i)}_{s_{\text{nocol}}} = P_{\text{nocol}_m} \cdot CP_m^{(i)}, \tag{B.16}$$

where the probability of capturing the channel is the coverage probability, $CP_m^{(i)}$, which can be determined as

$$\begin{aligned} CP_m^{(i)} &= \Pr(\gamma_m^{(i)} > \Gamma_m) = \Pr(c|h_i|^2 r_i^{-\alpha} \geq \Gamma_m) \\ &= \int_{l_{m-1}}^{l_m} \exp(-\frac{\Gamma_m r_i^{\alpha}}{c}) g_m(r_i) dr_i. \end{aligned} \tag{B.17}$$

Eq. (B.17) can be derived easily based on $|h_i|^2$ being exponentially distributed with average unit power and the PDF of the population to r_i being $g_m(r_i)$.

$k_{\mathbf{int},m} = 0$

In the case that there is solely co-SF interference, we can express the probability of capturing the channel as

$$\begin{aligned} CP^{(i)}_{\mathrm{co},m}(k_{\mathrm{co},m}) &= \Pr\left(\frac{|h_i|^2 r_i^{-\alpha}}{\sum_{k=1}^{k_{\mathrm{co},m}} |h_k|^2 r_k^{-\alpha} + \frac{1}{c}} \geq \Gamma_{\mathrm{co}} | k_{\mathrm{co},m}\right) \\ &\int_{l_{m-1}}^{l_m} \exp\left(-\frac{\Gamma_{\mathrm{co}} r_i^{\alpha}}{c}\right) g_m(r_i) [I(r_i)]^{k_{\mathrm{co},m}} dr_i, \end{aligned} \tag{B.18}$$

where $I(r_i)$ is equal to

$$I(r_i) = \int_{l_{m-1}}^{l_m} \frac{1}{1 + \Gamma_{\mathrm{co}} (\frac{r_i}{r})^{\alpha}} g_m(r) dr. \tag{B.19}$$

$k_{\mathbf{co},m} = 0$

In this case, there is no co-SF interference, $CP^{(i)}_{\mathrm{int},m}(k_{\mathrm{int},m})$ can be expressed as

$$\begin{aligned} CP^{(i)}_{\mathrm{int},m}(k_{\mathrm{int},m}) &= \Pr\left(\frac{|h_i|^2 r_i^{-\alpha}}{\sum_{k=1}^{k_{\mathrm{int},m}} |h_k|^2 r_k^{-\alpha} + \frac{1}{c}} \geq \Gamma_{\mathrm{int},m} | k_{\mathrm{int},m}\right) \\ &\int_{l_{m-1}}^{l_m} \exp\left(-\frac{\Gamma_{\mathrm{int},m} r_i^{\alpha}}{c}\right) g_m(r_i) [\tilde{I}(r_i)]^{k_{\mathrm{int},m}} dr_i, \end{aligned} \tag{B.20}$$

$\tilde{I}(r_i)$ is defined as

$$\tilde{I}(r_i) = \int_{R \setminus R_m} \frac{1}{1 + \Gamma_{\mathrm{int},m} (\frac{r_i}{r})^{\alpha}} g_p(r) dr, \tag{B.21}$$

$k_{\mathbf{co},m} \cdot k_{\mathbf{int},m} \neq 0$

In case both co-SF and inter-SF interference are present. $CP^{(i)}_{\mathrm{int},m}((k_{\mathrm{int},m}))$ can be expressed as

$$\begin{aligned} &CP^{(i)}_{\mathrm{co,int},m}(k_{\mathrm{co},m}, k_{\mathrm{int},m}) = \\ &pr\left(\frac{|h_i|^2 r_i^{-\alpha}}{\sum_{k=1}^{k_{\mathrm{co},m}} |h_k|^2 r_k^{-\alpha} + \sum_{j=1}^{k_{\mathrm{int},m}} |h_j|^2 r_j^{-\alpha} + \frac{1}{c}} \right. \\ &\left. \geq \max(\Gamma_{\mathrm{co}}, \Gamma_{\mathrm{int},m}) | k_{\mathrm{int},m}, k_{\mathrm{co},m} \right) \\ &= \int_{l_{m-1}}^{l_m} \exp\left(-\frac{\Gamma_{\mathrm{co}} r_i^{\alpha}}{c}\right) g_m(r_i) [I(r_i)]^{k_{\mathrm{co},m}} [I'(r_i)]^{k_{\mathrm{int},m}} dr_i, \end{aligned} \tag{B.22}$$

where $\max(\Gamma_{\text{co}}, \Gamma_{\text{int},m})$ is always equal to Γ_{co}. $I'(r_i)$ is defined as

$$I'(r_i) = \int_{R \setminus R_m} \frac{1}{1 + \Gamma_{\text{co}}(\frac{r_i}{r})^{\alpha}} g_p(r) dr. \tag{B.23}$$

Then the probability of capturing the channel can be expressed as a weighed sum of the derived capture probabilities, where the weights are the probabilities of the particular capture scenario taking place.

$$\begin{aligned} P_{\text{cap}_m} &= P_{\text{nocol}_m} \cdot CP_m^{(i)} \\ &+ \sum_{k_{\text{co},m}=1}^{\infty} P_{(k_{\text{int},m}=0)} \cdot P_{k_{\text{co},m}} \cdot CP_{\text{co},m}^{(i)}(k_{\text{co},m}) \\ &+ \sum_{k_{\text{int},m}=1}^{\infty} P_{(k_{\text{co},m}=0)} \cdot P_{k_{\text{int},m}} \cdot CP_{\text{int},m}^{(i)}(k_{\text{int},m}) \\ &+ \sum_{k_{\text{int},m}=1}^{\infty} \sum_{k_{\text{co},m}=1}^{\infty} P_{k_{\text{co},m}} \cdot P_{k_{\text{int},m}} \cdot CP_{\text{co,int},m}^{(i)}(k_{\text{co},m}, k_{\text{int},m}) \end{aligned} \tag{B.24}$$

Then, $FCP_m^{(i)}$ is given by

$$FCP_m^{(i)} = P_{\text{cap}_m}(\tau = T_{\text{f}_m}). \tag{B.25}$$

3.2 Calculation of $FDP^{(i)}$ and Throughput

A SX1276 based LoRa receiver will 'lock onto' a frame once it has received 4 preambles as supported by [7, 8]. Preamble detection and frame demodulation are separated in the SX1301 and while it is capable of demodulating 8 frames simultaneously as supported by [11, 12], it is able to detect 48 preambles at once, i.e. a preamble for every SF and channel combination.

Since the preamble symbols and the 4 concurrent symbols needed for detection only constitute a small fraction of the total frame size, we approximate the capture probability as the coverage probability.

$$CP_{\text{pre}_m}^{(i)} \approx CP_m^{(i)}. \tag{B.26}$$

We evaluate the FDP using the CDF of the Poisson distributed number of frames received by the demodulator at any given instant, $k_M = \sum_{m=7}^{12} k_m$ where k_m denotes

the number of frames being received using m such that
$L_M = \sum_{m=7}^{12} L_m$ where $L_m = \lambda \cdot \delta_m \cdot T_{\mathrm{f}_m} \cdot CP^{(i)}_{pre_m} \cdot (1 - FDP^{(i)})$

$$FDP^{(i)} = 1 - \exp(-L_M) \cdot \sum_{k=0}^{7} \frac{(L_M)^k}{k!} \tag{B.27}$$

Eq. (B.27) can be interpreted as the probability of there not being 7 or less concurrent frame receptions at the beginning of the reception of a new frame. $FSP^{(i)}_m$ is now computable as Eq. (B.12) by using Eq. (B.25) and Eq. (B.27).

The throughput of the cell can then be calculated as

$$T_{\Delta_{\mathrm{Scheme}}} = \sum_{m \in M} FSP^{(i)}_m \cdot \delta_m \cdot \lambda \cdot B \tag{B.28}$$

where the throughput is defined as successfully received bytes of LoRaWAN payload per second.

3.3 SF Allocation

In this subsection, we compute δ_m, l_m and $g_m(r)$ for $m \in M$ for three SF allocation schemes. The cell center and edge for every scheme are defined as $l_6 = 0$ and $l_{12} = R$. All parameters for Δ which are presented in this subsection, can be found in Tab. B.2.

Uniform

In this scheme, the cell is not divided into annuli and instead SFs are assigned to devices uniformly. We denote the SF allocation set for this scheme by Δ_{Uni}. Thus, $\delta_m = \frac{1}{6}$ and $g_m(r) = g_R(r)$. In this case, $l_m = R$ and $l_{m-1} = 0$ for $m \in M$ is used in the model.

Distance

We denote the SF allocation set for this scheme by Δ_{Dist}. We compute the borders of each annulus based on the deterministic path loss and the receiver sensitivity such that $l_m = \left(\frac{P_0 \cdot A(f_c)}{\theta_{\mathrm{RX}_m}} \right)^{\frac{1}{\alpha}}$. Then, we calculate δ_m as a function of the area of a given annulus to the total cell area $\delta_m = \frac{N_m}{N} = \frac{A_m}{A}$. N_m is the number of devices that are assigned SF m and A_m denotes the area of the annuli m. This means that $g_m(r) = \frac{g_R(r)}{\delta_m}$.

Table B.2: ©2019 IEEE: SF Allocation schemes

SF	Δ_{Uni}		Δ_{Dist}		Δ_{Eqload}	
	δ_m	l_m	δ_m	l_m	δ_m	l_m
7	1/6	-	.2	.45	.47	.45
8	1/6	-	.08	.54	.25	.54
9	1/6	-	.11	.64	.14	.64
10	1/6	-	.17	.76	.08	.76
11	1/6	-	.19	.88	.04	.88
12	1/6	-	.25	1	.02	1

Equivalent load

We let the l_m be the same as for Δ_{Dist}. In this allocation scheme, denoted Δ_{Eqload}, we calculate T_{f_m} for each SF and assign SFs to keep the load equal for each SF such that $\delta_m = \frac{1/T_{\text{f}_m}}{\sum_{s=M} 1/T_{\text{f}_s}}$. We keep the device density uniform throughout each annulus, but we assume a higher device density in the annuli closer to the gateway such that $\sum_{m \in M}(\delta_m | l_m) = 1$.

4 Results and discussion

We have evaluated CP, FCP, FDP and FSP analytically using Eq. (B.17), Eq. (B.25), Eq. (B.27) and Eq. (B.12), which are derived in Sec. 3. Furthermore, a Matlab-based simulation of the unacknowledged LoRaWAN uplink was implemented and results were simulated. In the simulation, the reception conditions 1-4 are evaluated at the symbol level, whereas the analytical approximation is evaluated at the frame-level due to (B.13) and (B.14). This results in the FCP approximation being lower-bound. SX1301 is capable of parallel reception on 8 LoRa channels, $1 \leq N_c \leq 8$. In this article we have assumed that the frequency of all N_c channels is the same for the evaluation of FCP, which is a fair assumption since all LoRaWAN channels in the EU 863-870MHz ISM Band are fairly close in frequency. The channels are still considered orthogonal to each-other when evaluating the FCP. The FDP takes into account traffic on all channels as discerned LoRa frames compete for the same demodulation paths. The FCPs of SF $\{7, 10, 12\}$ are plotted in Fig. B.3 for Δ_{Dist}. Calculating the load contribution on the FDP for each SF m, L_m/L_M, we get .59 for SF 12, .1 for SF 10 and .015 for SF 7. Although calculated for a specific allocation scheme, we can make some general assertions with regards to the impact of different SFs on the reception probability. The higher SFs take up demodulations paths for much longer time, which makes any transmission for SF $m \in M$ more likely to be dropped. The FCP is also lower for the higher SF frames, which have a greater chance of experiencing interference. A higher

coverage probability is seen for SF 7, which can be explained by SF 7 being allocated to devices in a punctured disc around the gateway.

In Fig. B.4 and Fig. B.5, we observe that the coverage probability for Δ_{Dist} and Δ_{Eqload} are .84 and .88, respectively. The slight increase in coverage probability from Δ_{Dist} to Δ_{Eqload} can be explained by the allocation of more devices in the SF 7 annulus. It is also evident that the FDP has a non-negligible effect on the FSP especially when more devices are allocated to higher SFs, which is the case for Δ_{Dist}.

Results for the CP, $FCP_m^{(i)}$ and $FSP_m^{(i)}$ are not plotted for the uniform allocation scheme because the frame-based evaluation of interference results in a very low bound for $FCP_m^{(i)}$. This is due to the mixture of frames of both high and low SFs in the uniform allocation scheme.

The throughput of the three allocation schemes are depicted Fig. B.6. While the Δ_{Dist} provides better coverage than Δ_{Uni}, it does not provide a much larger throughput on a cell basis since many devices are assigned to higher SFs. The equivalent-load allocation scheme assumes an uneven distribution of devices, which may not be the case often in actual deployments, but we see that the throughput is much larger in this case. This hints to cooperative assignment of SFs between gateways may possibly provide remarkable throughput improvements in LoRaWAN.

5 Conclusion

We have evaluated the uplink performance of a single gateway LoRaWAN deployment in terms of coverage, frame reception success probability and throughput. Collisions are evaluated in the time domain based on a traffic model in contrast to other works on capture effect in LoRaWAN, which evaluate capture effect for a fixed number of concurrent transmissions. Unlike previous works, we have considered the demodulation capabilities of the gateway and specifically evaluated SX1301, although the model is applicable to any chipset. We showed that the demodulation capabilities of the receiver have a large impact on the probability of receiving frames successfully.

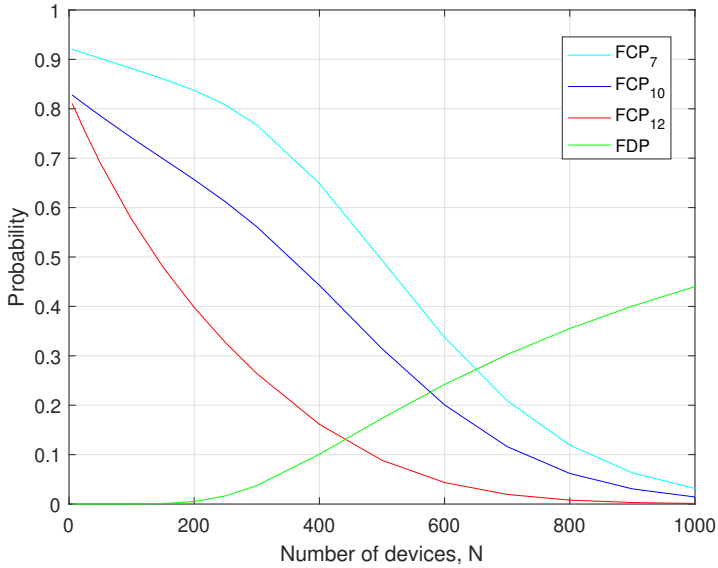


Fig. B.3: ©2019 IEEE: Contribution of SFs 7, 10 and 12 on frame capture probability and frame drop probability for Δ_{Dist}. The considered traffic model was a payload size of 50 Bytes and a mean inter-frame generation time of ten minute for each device. The traffic model is assumed to be the same for the 8 channels, which are considered.

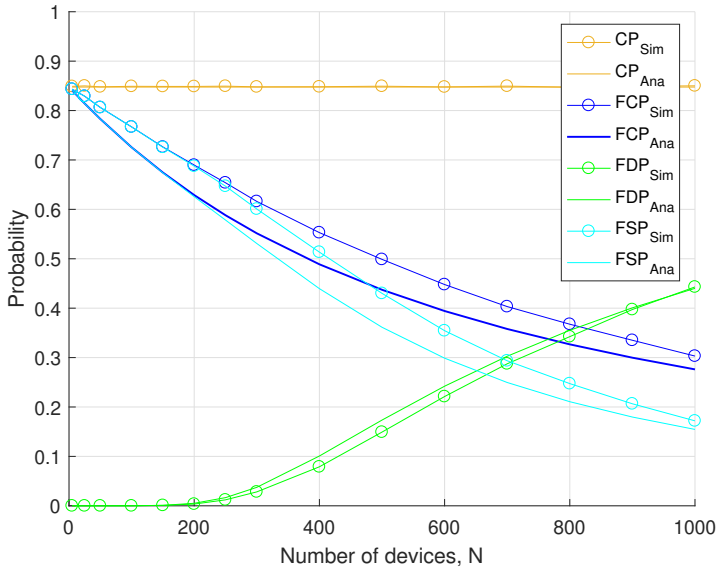


Fig. B.4: ©2019 IEEE: Probabilities associated with frame capture for Δ_{Dist}.
$\lambda_d = 1/600$, $B = 50$ and $N_c = 8$.

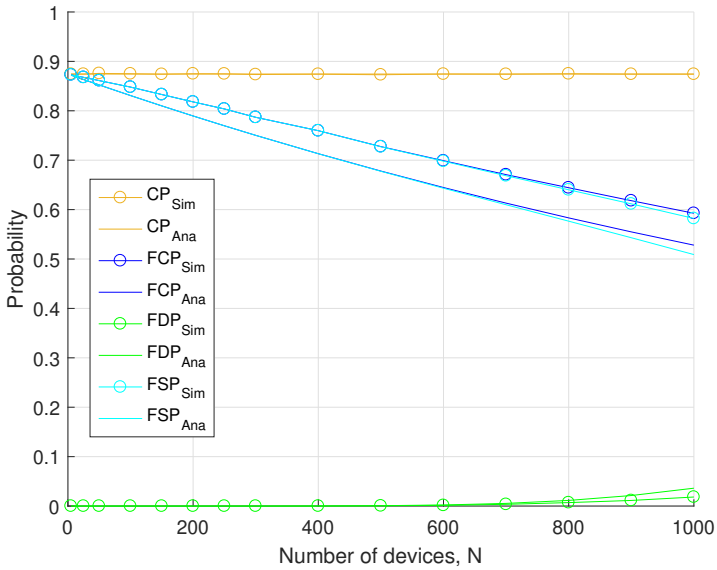


Fig. B.5: ©2019 IEEE: Probabilities associated with frame capture for Δ_{Eqload}. $\lambda_d = 1/600$, $B = 50$ and $N_c = 8$.

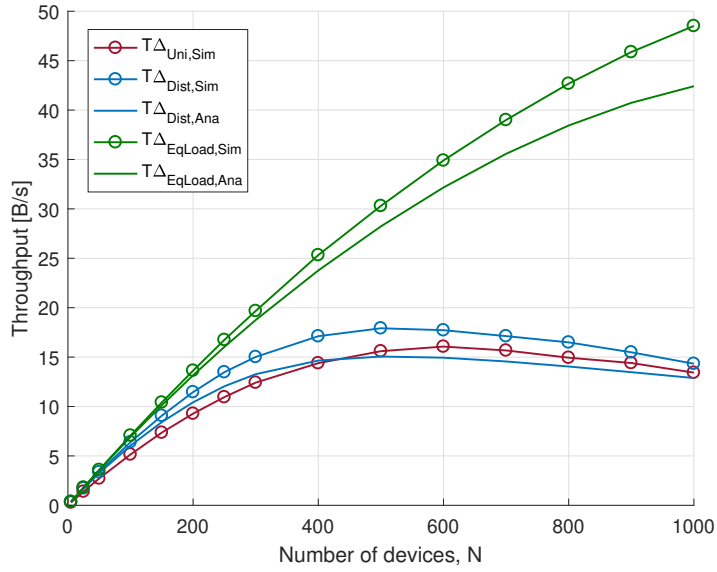


Fig. B.6: ©2019 IEEE: ©IEEE: Throughput for the LoRaWAN payload [B/s]. $\lambda_d = 1/600$, $B = 50$ and $N_c = 8$.

References

[1] F. Adelantado, X. Vilajosana, P. Tuset-Peiro, B. Martinez, J. Melia-Segui, and T. Watteyne, "Understanding the limits of lorawan," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 34–40, 2017.

[2] O. Georgiou and U. Raza, "Low power wide area network analysis: Can lora scale?" *IEEE Wireless Communications Letters*, vol. 6, no. 2, pp. 162–165, April 2017.

[3] Z. Li, S. Zozor, J. Drossier, N. Varsier, and Q. Lampin, "2d time-frequency interference modelling using stochastic geometry for performance evaluation in low-power wide-area networks," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.

[4] J. Lim and Y. Han, "Spreading factor allocation for massive connectivity in lora systems," *IEEE Communications Letters*, vol. 22, no. 4, pp. 800–803, April 2018.

[5] A. Mahmood, E. G. Sisinni, L. Guntupalli, R. Rondon, S. A. Hassan, and M. Gidlund, "Scalability analysis of a lora network under imperfect orthogonality," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2018.

[6] A. Waret, M. Kaneko, A. Guitton, and N. E. Rachkidy, "Lora throughput analysis with imperfect spreading factor orthogonality," *IEEE Wireless Communications Letters*, pp. 1–1, 2018.

[7] A. Rahmadhani and F. Kuipers, "When lorawan frames collide," *Proc. of the 12th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (ACM WiNTECH 2018)*, 2018.

[8] Semtech, *SX1276/77/78/79 Datasheet*, rev. 5 ed., Aug. 2016.

[9] L. A. T. Committee, *LoRaWAN Specification*, v1.1 ed., LoRa Alliance, Oct. 2017.

[10] R. B. Sørensen, D. M. Kim, J. J. Nielsen, and P. Popovski, "Analysis of latency and mac-layer performance for class a lorawan," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 566–569, Oct 2017.

[11] Semtech, *SX1301 Datasheet*, v2.4 ed., June 2017.

[12] ——. (2017) Demodulation capacity of the sx1301. (Visited: 2018-9-26). [Online]. Available: https://semtech.force.com/lora/articles/FAQ/Demodulation-capacity-of-the-SX1301.

Paper C

On Symbol-wise Collisions and Demodulation Path Blocking in Multi-Gateway LoRaWAN

René Brandborg Sørensen, Nazrin Razmi, Jimmy Jessen Nielsen, Petar Popovski

The paper has been **submitted, 2020**.

The layout has been revised.

Abstract

LoRaWAN has received much attention from industry and academia as a candidate for low-cost LPWAN connectivity. Thus the performance of LoRaWAN has been evaluated for various capture and collision models, but so far all works have evaluated frame reception failure against the average number of collisions or average SNR over the entire frame duration. In this letter, we consider individual symbol failures due to collisions and develop an approximation of the distribution of number of maximal number of colliding frames within a any symbol of the frame. We use this novel approach to analyze frame reception in multi-gateway LoRaWAN under an important practical limitation posed by the finite number of demodulation paths. The results show that in the case of a collision model, increasing the number of gateways yields better performance in terms of outage for individual devices, but with exponentially diminishing diversity gains. The exponent by which the diversity gains diminishes is proportional to the number of end devices.

1 Introduction

An important enabler of Internet of Things (IoT) is the technology of Low Power Wide Area Networks (LPWANs), and in particular Long Range Wide Area Network (LoRaWAN). LoRaWAN is a MAC protocol based on the LoRa PHY-layer, which modulates messages in chirps with different frequency-time gradients depending on Spreading Factors (SFs). LoRaWAN provides connectivity for end devices (EDs) through gateways (GWs) deployed in a star-topology as depicted in Fig. C.1. The RAN relays messages to network servers that collect received messages from multiple GWs and combine them to recreate the message history of each individual ED.

LoRaWAN's performance has been evaluated for various capture and collision models, see [1–4]. The authors in [1], provide a framework to evaluate co-SF interference by considering the level of time-frequency overlap of the interfering packets. In [2] and [3] the performance of LoRaWAN was evaluated by taking capture effect into account, which has been expanded upon in [3, 5] where the non-orthogonality between SFs was taken into account. In [4] a capture model for multiple simultaneous frames on a single SF LoRaWAN is presented. The performance given a multiple gateway deployment has been evaluated in [6–8] using stochastic geometry for various capture and interference conditions. Packet reception in real-world, multiple GW deployment has been evaluated experimentally in [9]. Interestingly, the maximal RSSI and SNR for a transmission was not found to be among the geographically closest GW-ED pairs, although closer GWs had, on average, better RSSI and received SNR. Also, multiple GWs were found to receive each transmission, sometimes even very remote GWs.

Notably, SIC has been evaluated for LoRa transceivers in [10] using stochastic geometry in combination with a capture model to realistically evaluate the potential gain of

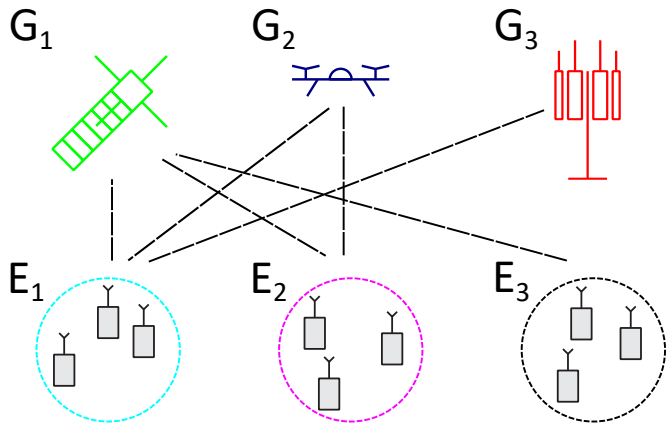


Fig. C.1: I resized the E and G, are they OK?An illustration of a graph network consisting of three groups of EDs connected to three different GWs. In this example G_1 is connected to all groups, G_2 to $E_1 \cup E_2$ and G_3 to E_1. E_1 is similarly connected to all GWs, E_2 to $G_1 \cup G_2$ and E_3 to G_1 only.

employing SIC for LoRa. However, SIC is not currently employed in Semtechs modems so in [11] the SX1276 LoRa transceiver chipset was found to drop both transmissions when the first has been detected and a second transmission causes capture during demodulation of the first transmission. Based on this timing-consideration, we presented a joint timing and capture effect model in [12] for the performance of LoRaWANs uplink in a single cell. The model was based on the SX1301 chipset, that is widely used in GWs.

The commonly used time-domain model in [12] and similar works finds the number of interfering transmissions over an entire frame. In this letter we generalize this model by catering for the asynchronism among the interfering frames and analyze the maximal number of ongoing interfering transmissions during a particular transmitted symbol. This new collision model can also be imported into models based on stochastic geometry to account for the capture effect. In this work we will focus on the a simple collision model for a graph network, ie. a set of nodes connected by stochastic links as in Fig. C.1, in order to investigate the benefits in terms of diversity gain from deploying multiple SX1301-based GWs, each operating under the constraint that the demodulation paths are reserved for decoding ongoing frames.

2 System model

We consider a deployment with n_{E} groups of Class A EDs and n_{G} GWs. A group of EDs E_i for $i \in \ I = \{1, 2, ..., n_{\mathrm{E}}\}$ is connected to a GW G_j for $j \in \ J = \{1, 2, ..., n_{\mathrm{G}}\}$ with a probabilistic adjacency matrix $\mathbf{P}$, whose elements $p_{i,j}$ are probabilities that a

given frame is exchanged successfully between an ED in E_i and G_j.

This scenario is sketched in Fig. C.1 for $n_{\mathrm{E}} = n_{\mathrm{G}} = 3$ and

$$\mathbf{P} = \begin{Bmatrix} .5 & .5 & .5 \\ .5 & .5 & 0 \\ .5 & 0 & 0 \end{Bmatrix} . \tag{C.1}$$

The EDs in E_i transmit at an aggregated rate λ_{E_i} yielding Markovian arrivals with the rate $\lambda_{\mathrm{G}_j} = \sum_{i \in I} p_{i,j} \lambda_{\mathrm{G}_i}$ at G_i. Each ED is assigned a fixed SF m to be used for transmission, where $M = \{7, 8, ..., 12\}$ and $m \in M$ as allowed in LoRaWAN [13]. Furthermore, the ED is assigned a specific channel ch where $ch \in C = \{0, 1, ..., c-1\}$ for a total of c channels, each assumed to have bandwidth $BW = 125$ [kHz]. The symbol period for a given SF m is $T^m_{\mathrm{Sym}} = \dfrac{2^m}{BW}$. The fraction of UEs that are allocated to SF m is denoted α_m. We consider a uniform distribution for the allocation of SFs, $\alpha_m = 1/6$, but the model is applicable to any SF allocation.

Each transmission is a LoRa frame with a payload of B bytes. A LoRa frame consists of a preamble that is η_{pre} at 8 symbols long in the EU868 ISM band, a LoRa header, payload and, optionally, a CRC for the payload. The number of symbols in the LoRa header and payload of a LoRa frame for SF m, η^m_{pay} is [14]:

$$\eta_{\mathrm{pay}_m} = 8 + \max\left(\left\lceil \frac{(8B - 4m + 44)}{(4m - 2 \cdot \mathbb{I}_{\mathrm{DE}})} \right\rceil \cdot (CR + 4), 0 \right), \tag{C.2}$$

where $CR = i$ for $i \in \{1, 2, 3, 4\}$ and corresponds to a coding rate of $\frac{4}{4+i}$. The indicator $\mathbb{I}_{\mathrm{DE}}$ is one if low data rate optimization is used and zero otherwise. Low data rate optimization is mandatory for SF $\{11, 12\}$ because it increases robustness towards clock drift. The total number of symbols in a LoRa frame for SF m is $\eta^m_{\mathrm{frm}} = \eta^m_{\mathrm{pre}} + \eta^m_{\mathrm{pay}}$ and the total transmission time is $T^m_{\mathrm{frm}} = T^m_{\mathrm{Sym}} \cdot \eta^m_{\mathrm{frm}}$. We assume that all EDs transmit at such a rate that duty cycling does not impact transmission times and furthermore that the aggregate transmission times are a Markovian process.

In order for a symbol to be received correctly, the signal-to-interference-ratio (SIR) for co-SF and inter-SF must be larger than the thresholds as given by [5], corresponding to the maximal number of allowable concurrent transmissions, $\theta_{m_{\mathrm{ref}}, m_{\mathrm{int}}}$, as listed in Tab. C.1. For example, $\theta_{12,7} = \theta_{12,8} = \theta_{12,9} = 316$ so an SF 12 modulated frame is lost if there are more than 316 concurrent interfering SF 7, SF 8, or SF 9 transmissions.

In order to successfully receive a LoRa frame, the GW must detect the preamble and demodulate both the header and payload [11]. A LoRa receiver chain is depicted in Fig. C.2 for the SX1301 chipset, which is designed and widely used in GWs. Sx1301 is capable of detecting 48 preambles at once, corresponding to each possible SF in 8 channels in parallel. The receiver only detects a transmission if $\eta_{\mathrm{con}} = 4$ consecutive symbols in the preamble are received correctly. If a preamble is detected, a demodulation path

$SF_{ref} \backslash SF_{int}$	7	8	9	10	11	12
7	0	6	7	7	7	7
8	12	0	12	15	19	19
9	31	19	0	19	25	31
10	79	63	50	0	50	63
11	158	158	125	99	0	99
12	316	316	316	251	199	0

Table C.1: The maximal number of concurrent transmissions, $\theta_{m_{\text{ref}},m_{\text{int}}}$, of m_{int} for successful transmission of m_{ref} given the SIRs for Sx1271 in [5] and the channel gain of 1 considered in our system model.

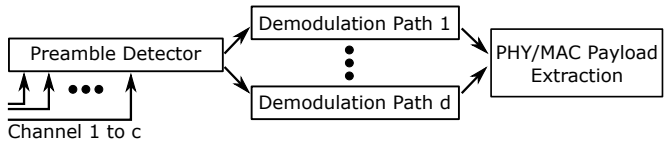


Fig. C.2: Reception chain for a LoRa transceiver. Sx1301 is capable of detecting 48 preambles at once, corresponding to each possible SF in 8 channels, and $d = c = 8$ [15].

will be assigned to the frame. SX1301 has 8 demodulation paths that allow parallel demodulation of frames across any combination of SFs and channels. Any detected excess concurrent frame is dropped.

3 Analysis

We divide a reference frame into η_{ref} reference symbol periods. Let the number of transmissions initiated in period j be a random variable denoted K_j. We enumerate j starting at $j = 1$ exactly one interfering frame period away from the beginning of a reference frame by defining an extended period where initiated transmissions will still be ongoing within the first symbol of the reference frame. The number of reference symbols in this extended period is $\eta_{\text{int}} = \left\lceil \frac{T_{\text{frm}}^{m_{\text{int}}}}{T_{\text{Sym}}^{m_{\text{ref}}}} \right\rceil$. This enumeration of symbol periods is depicted in Fig. C.3.

The number of ongoing transmissions within symbol i of the reference frame is then a random variable denoted by S_i, is the sum of the transmissions initiated within the period of an interfering frame.

$$S_i = \sum_{j=i}^{i+\eta_{\text{int}}} K_j \ . \tag{C.3}$$

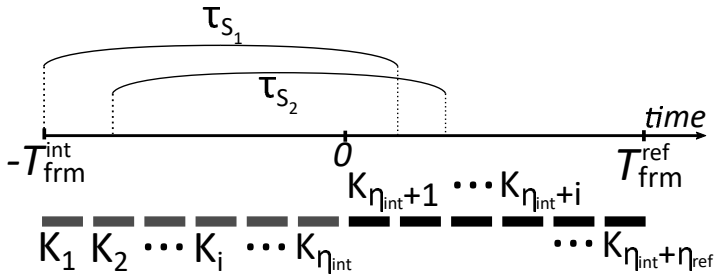


Fig. C.3: Representation of a reference frame and an extended interference period as a set of symbol periods.

The total number of ongoing transmissions within the period of the reference frame is denoted by F such that $F = \sum_{j=1}^{\eta_{\text{int}}+\eta_{\text{ref}}} K_j$.

Then we have that

$$\begin{aligned}\Pr(K_i = k_i) &= \text{Pois}(k_i, \lambda_k) \ , \\ \Pr(S_1 = s_1) &= \text{Pois}(s_1, \lambda_k(\eta_{\text{int}} + 1)) \ , \\ \Pr(F = f) &= \text{Pois}(f, \lambda_k(\eta_{\text{int}} + \eta_{\text{ref}})) \ ,\end{aligned}$$

where λ_K is the arrival rate within each reference symbol period $\lambda_K = T_{\text{Sym}}^m \lambda_{\text{G}_i} \alpha_m / c$ and τ_S is the period that initiating interfering transmissions will interfere with the reference symbol $\tau_S = T_{\text{Sym}}^{m_{\text{ref}}}(\eta_{\text{int}} + 1)$. We also point out that for Poisson distributions we have that

$$\Pr(S_1 - K_1 = s_1 - k_1) = \text{Pois}(s_1 - k_1, \lambda_k(\eta_{\text{int}})) \ ,$$

First we describe the probability distribution of the maximum number of ongoing transmissions within one symbol period. It is clear that, $S_{i+1} = S_i - K_i + K_{i+\eta_{\text{int}}}$ so to find the CMF for each subsequent time-slot we must consider probability distributions that are conditioned on the number of ongoing transmissions in the previous time-slot.

Since S_i and K_i are Poisson distributed we can find K_i conditioned on S_i easily as

$$\Pr(K_i = k_i | S_i = s_i) = \frac{\Pr(K_i = k_i) \Pr(S_i - K_i = s_i - k_i)}{\Pr(S_i = s_i)} \tag{C.4}$$

Then similarly, since $s_{i+1} = s_i - k_i + k_j$ we have that

$$\begin{aligned}&\Pr(S_{i+1} = s_{i+1} | S_i = s_i) \\ &= \sum_{s_i=0}^{\infty} \left(\frac{\Pr(K_i = k_i) \Pr(S_i - K_i = S_i - k_i) \Pr(K_j = k_j)}{\Pr(S_i = s_i)} \right)\end{aligned} \tag{C.5}$$

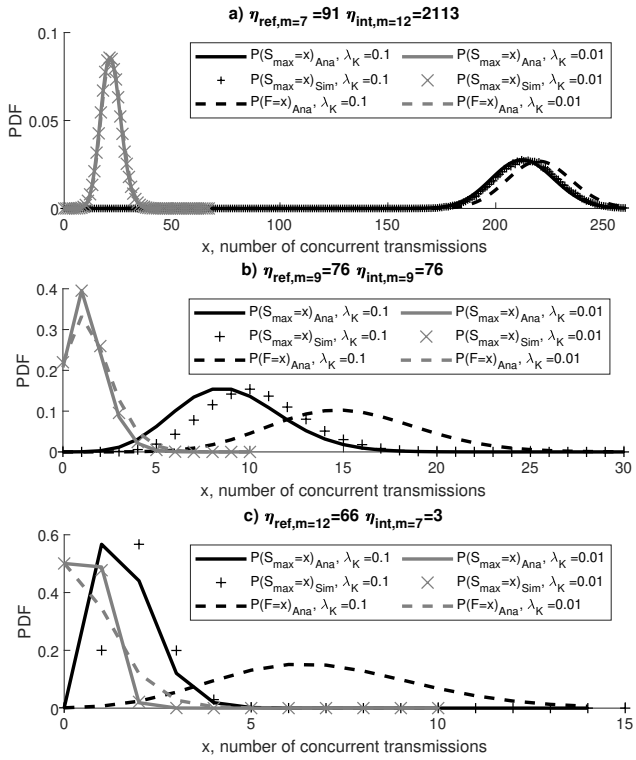


Fig. C.4: Approximation accuracy for the maximum number of concurrent transmissions in a single symbol for $\lambda_K = 0.1$ and $\lambda_K = 0.01$, respectively.

where $j = i+\eta_{\text{int}}$. We evaluate (C.5) numerically and found heuristically that a good approximation of the maximum number of ongoing transmissions during any one reference symbol, when $\lambda_K << 1$ is

$$\Pr(S_{\max} = x) \approx Pr(S_1 = x) \Pr(S_{i+1} \leq s_{i+1} | S_i = s_i)^{\eta_{\text{ref}}} . \tag{C.6}$$

Another heuristic that gives a good approximation when $\lambda_K \approx 1$ is

$$\Pr(S_{\max} = x) \approx$$
$$\Pr(S_i \leq x)^{\frac{\eta_{\text{ref}}}{\eta_{\text{int}}+1}} \Pr(S_1 = x) \Pr(S_2 \leq x | S_1)^{\eta_{\text{ref}} - \frac{\eta_{\text{ref}}}{\eta_{\text{int}}+1}} .$$

In our scenario $\lambda_K << 1$, so we use (C.6). The accuracy of this approximation is depicted in Fig. C.4 along with a comparison to $\Pr(F = f)$, which is the distribution

naively used to evaluate the number of collisions within the frame. It is clear that $\Pr(F = f)$ is a poor approximation of $\Pr(S_{\max} = x)$, whereas the presented model is a close approximate.

Now, we can evaluate the probability that the payload is demodulated correctly as $Pr(S_{\max} \leq \theta_{m_{\text{ref}},m_{\text{int}}})$, where $S_{\max}$ is evaluated for the arrival rate of the interfering SF m_{int} within a symbol period of the reference SF m_{ref}. Then we can find the probability of not failing due to collisions at GW G_j

$$\Pr(\text{G}^j_{\text{PayOK}}) = \sum_{i \in M} \alpha_{m_i} \prod_{j \in M} \text{PayOK}_{m_i,m_j} \ . \tag{C.7}$$

The preamble is relatively small, so we can in good faith assume that the probability of receiving the preamble correctly is equivalent to the probability of receiving a single symbol correctly as before

$$\Pr(\text{G}^j_{\text{PreOK}}) = \sum_{i \in M} \alpha_{m_i} \prod_{j \in M} \Pr(S_1 \leq \theta_{m_{\text{ref}},m_{\text{int}}}) \ . \tag{C.8}$$

We make the assumption that frames are received at the GW as a Markovian process, because a large number of independent EDs are transmitting in an uncoordinated fashion. Then the well-known formula for Erlang-B can be applied in order to compute the blocking rate

$$\Pr(\text{G}^j_{\text{DemOK}}) = 1 - \frac{\frac{\lambda^d_{\text{DP}}}{d!}}{\sum_{x=1}^{d} \frac{\lambda^x_{\text{DP}}}{x!}} \ , \tag{C.9}$$

where the arrivals at the demodulation paths is the arrivals for which preambles are detected, so the arrival rate is $\lambda_{\text{DP}} = \sum_{m \in M} (\lambda_d T^m_{\text{frm}} n_{\text{devs}} \alpha_m \Pr(\text{PreOK}_{m_i}))$ and $d = 8$ for the SX1301 chipset.

In order for a frame to be received successfully the preamble must be detected, a demodulation path must be available, and the payload must be received correctly. We assume, $\Pr(\text{PreOK}^i | \text{PayOK}^i = 1) \approx 1$ for a transmission i, since the preamble is small, so the probability of outage for a transmission in G^j is

$$\Pr(\text{G}^j_{\text{Outage}}) = 1 - \Pr(\text{DemOK}) \Pr(\text{PayOK}) \ . \tag{C.10}$$

If we make the assumption that frame receptions at all GWs are independent, then

the factors of outage from the perspective of a UE in E_i is

$$\Pr(E^i_{\text{PreOK}}) = 1 - \prod_{j \in J} (1 - p_{i,j} \cdot (1 - \Pr(G^j_{\text{PreOK}}))) \,, \tag{C.11}$$

$$\Pr(E^i_{\text{PayOK}}) = 1 - \prod_{j \in J} (1 - p_{i,j} \cdot (1 - \Pr(G^j_{\text{PayOK}}))) \,, \tag{C.12}$$

$$\Pr(E^i_{\text{DemOK}}) = 1 - \prod_{j \in J} (1 - p_{i,j} \cdot (1 - \Pr(G^j_{\text{DemOK}}))) \,, \tag{C.13}$$

$$\Pr(E^i_{\text{Outage}}) = \prod_{j \in J} (1 - p_{i,j} \cdot \Pr(G^j_{\text{Outage}})) \,. \tag{C.14}$$

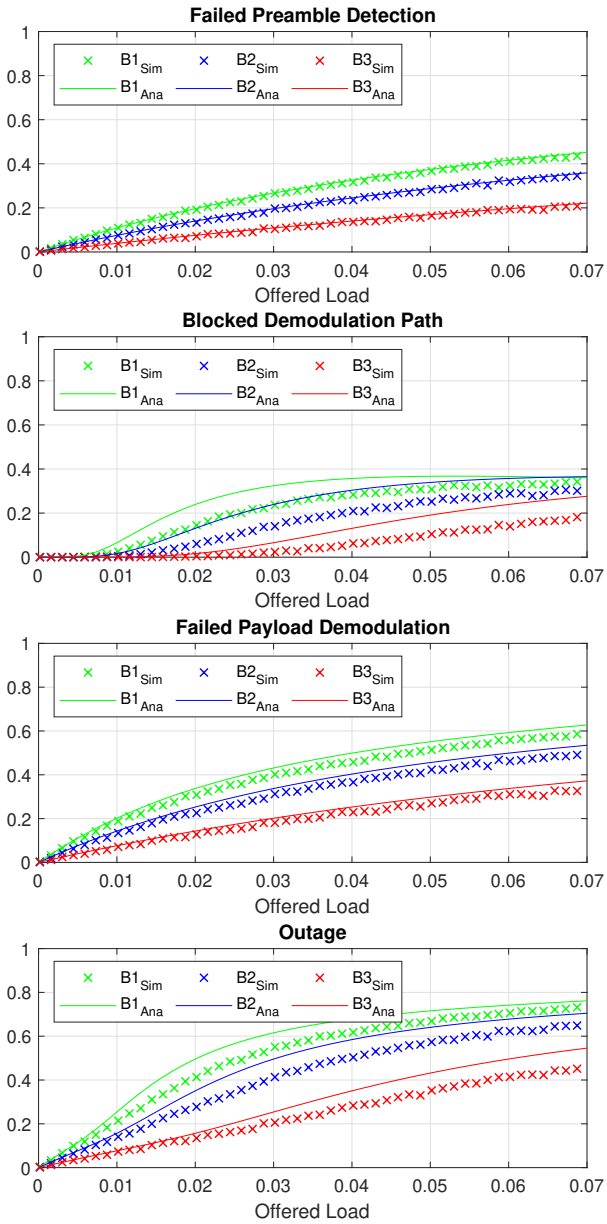
4 Results

In addition to the analytical results we have simulated the scenario in Fig. C.1 using the parameters given in Sec. 2. Note that the number of EDs in each ED group is set to 500, so a total of 1500 EDs are connected to G_1, 1000 EDs are connected to to G_2 and 500 EDs are connected with G_3.

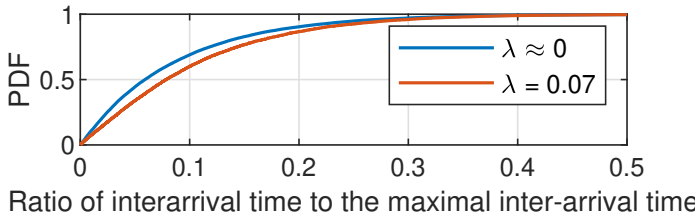
The outage factors as seen from the perspective of each GW is plotted in Fig. C.5a. Outage increases with the number of EDs. Notice that although failed payload demodulation is a major factor, failures at all stages in the receiver chain contribute in a non-negligible way to the overall outage. The analytical results are coherent with simulation results with the exception of the demodulation path blocking rate, which is lower in simulations than it should be for an Erlang-B queue. It is clear that the Erlang B formula is correct, so necessarily the assumption of Markovian arrivals at the demodulation paths of the receiver chain must be incorrect. Upon further investigation we found that collisions in the preamble detection shifts the inter-arrival distribution towards longer inter-arrival times as shown for the distribution of simulated values in Fig. C.6a for G_1. So the Markovian assumption of time-invariant arrivals does not hold, because arrivals at the demodulation paths are likely to be further apart in time. Thus the Erlang B formula appears to be an upper bound.

In Fig. C.5a we also see that using F instead of $S_{\max}$ yields a food approximation for G^j_{PayOK} when $\lambda_K << 1$. This is an effect of the adopted collision model where collision failures almost exclusively happens for same SF transmissions when $S_{\max} \geq 1$. which is $\Pr(S_{\max} = 0) = \Pr(F = 0)$. The difference in the distributions of F and $S_{\max}$ would however not be negligible for capture models.

The outage factors as perceived by a ED in group E_1, E_2 and E_3 is plotted in Fig. C.7a. Here, the analytical result for the demodulation path blockage is too high due to the discrepancy in the analytical blocking rate for each GW, but also due to the assumed independence in blocking between the gateways.



(a) Quality of service as seen from a GW's perspective.



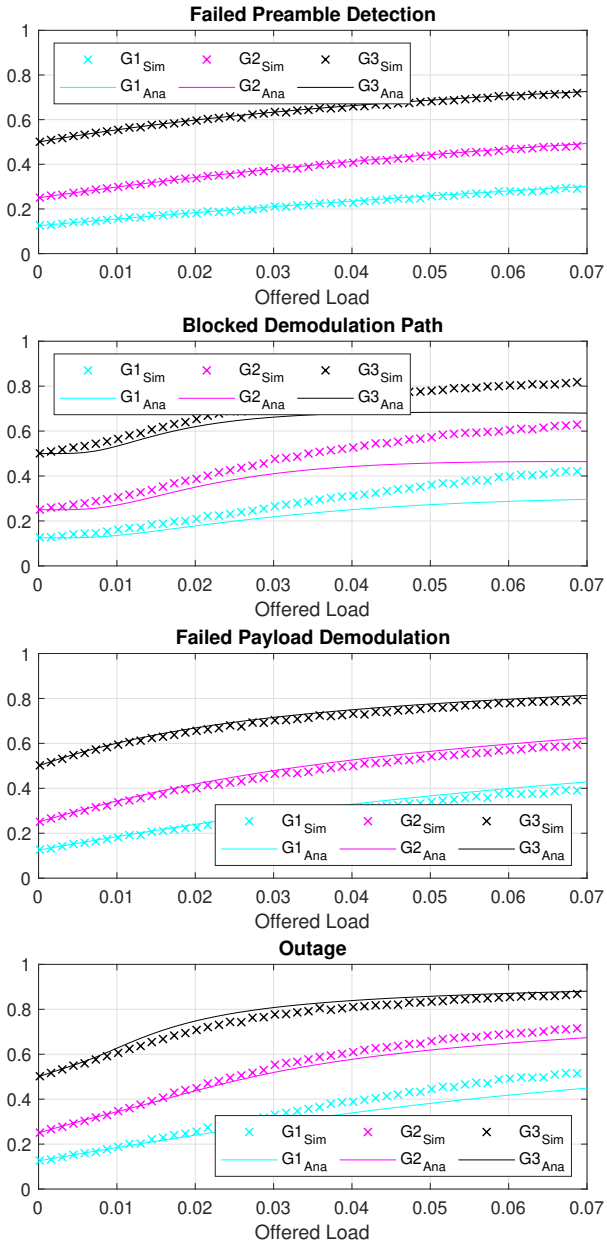
(a) Inter-arrival distribution shift in B_1 with increasing traffic. At higher traffic intensity, it becomes non-Markovian as short-inter arrival times are likely to experience collisions.

EDs see a diversity gain in performance from being connected to more GWs, although with a diminishing return. We can also observe this in Fig. C.7b that depicts the mean outage as seen by a ED in a group of 500 EDs for an increasing amount of GWs.

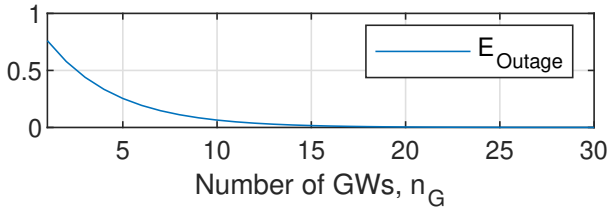
5 Conclusion

In this work we modelled the probability distribution of the maximal number of ongoing transmissions in any symbol, instead of the number of ongoing transmissions over an entire frame as is the conventional approach. We applied the model on a graph-network with a simple collision model for LoRaWAN.

Our analytical results fit well with simulation results and show as one would expect that given a fixed number of EDs we can add a sufficient number of GWs in order to provide any level of reception probability for an ED. However, the gain of each additional GW is diminishing. As the number of EDs is increased the diminishing effect is increased as a result of both an increase in collisions experienced by the GWs and increased blocking of the demodulation paths in the GWs.



(a) Quality of service as seen from a device's perspective.



(b) Device outage as a function of GW number for 500 connected devices.

References

[1] Z. Li, S. Zozor, J. Drossier, N. Varsier, and Q. Lampin, "2d time-frequency interference modelling using stochastic geometry for performance evaluation in low-power wide-area networks," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.

[2] A. Mahmood, E. G. Sisinni, L. Guntupalli, R. Rondon, S. A. Hassan, and M. Gidlund, "Scalability analysis of a lora network under imperfect orthogonality," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2018.

[3] A. Waret, M. Kaneko, A. Guitton, and N. E. Rachkidy, "Lora throughput analysis with imperfect spreading factor orthogonality," *IEEE Wireless Communications Letters*, pp. 1–1, 2018.

[4] A. Furtado, J. Pacheco, and R. Oliveira, "Phy/mac uplink performance of lora class a networks," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6528–6538, 2020.

[5] D. Croce, M. Gucciardo, S. Mangione, G. Santaromita, and I. Tinnirello, "Impact of lora imperfect orthogonality: Analysis of link-level performance," *IEEE Communications Letters*, vol. 22, no. 4, pp. 796–799, April 2018.

[6] Z. Qin, Y. Liu, G. Y. Li, and J. A. McCann, "Modelling and analysis of low-power wide-area networks," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–7.

[7] L. Beltramelli, A. Mahmood, M. Gidlund, P. Österberg, and U. Jennehag, "Interference modelling in a multi-cell lora system," in *2018 14th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct 2018, pp. 1–8.

[8] M. Ni, M. Jafarizadeh, and R. Zheng, "On the effect of multi-packet reception on redundant gateways in lorawans," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–6.

[9] K. Mikhaylov, M. Stusek, P. Masek, R. Fujdiak, R. Mozny, S. Andreev, and J. Hosek, "On the performance of multi-gateway lorawan deployments: An experimental study," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[10] J. M. d. S. Sant'Ana, A. Hoeller, R. D. Souza, H. Alves, and S. Montejo-Sánchez, "Lora performance analysis with superposed signal decoding," *IEEE Wireless Communications Letters*, pp. 1–1, 2020.

[11] A. Rahmadhani and F. Kuipers, "When lorawan frames collide," *Proc. of the 12th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (ACM WiNTECH 2018)*, 2018.

[12] R. B. Sørensen, N. Razmi, J. J. Nielsen, and P. Popovski, "Analysis of lorawan uplink with multiple demodulating paths and capture effect," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–6.

[13] LoRa Alliance Technical Committee, *LoRaWAN Specification*, v1.1 ed., LoRa Alliance, Oct. 2017.

[14] Semtech, *SX1276/77/78/79 Datasheet*, rev. 5 ed., Aug. 2016.

[15] ——, *SX1301 Datasheet*, v2.4 ed., June 2017.

Paper D

On multi server queues with degenerate service time distributions and no waiting lines (G/D/n/n)

René Brandborg Sørensen, Jimmy Jessen Nielsen, Petar Popovski

The paper has been **submitted, 2020**.

The layout has been revised.

Abstract

The growth of Machine-Type Communication (MTC) increases the relevance of queuing scenarios with deterministic service times. In this letter, we present a model for queues without waiting lines and with degenerate service time distributions and show how the framework is extendable to model general service time distributions. Simple bounds and a close approximation of the blocking probability are derived and the results are shown to hold for simulated queues with Markovian and degenerate arrival processes.

1 Introduction

A number of emerging use cases in Machine Type Communication (MTC) [1] require queueing models that are significantly different from traditional models used in teletraffic theory. A distinctive feature of many MTC applications is that machines are likely to perform different tasks within an almost-deterministic time. This brings relevance to queues with degenerate service times and non-Markovian, in particular periodical, arrival times. While networks in general have become packet-switched, relying heavily on buffering, there are still switching operations within communications that require immediate service or service within a very short time. Let us, for example, take the case of LoRa, where messages are modulated with one of seven different spreading factors. In LoRaWAN six spreading factors are used to create six quasi-orthogonal sub-channels in each channel within the network. LoRa gateway transceiver chipsets are capable of detecting preambles for every spreading factor simultaneously on multiple channels, but only a finite amount of demodulation paths are available [2, 3]. So messages in excess of the available demodulation paths are lost. Other general examples in telecommunications include: service of critical real-time interrupts [4], scheduling of immediate resources in FDMA networks and packet demodulation in FDMA networks.

We use Kendall's notation [5], noting that $\sim/\sim/n/n$ refers to queues of finite capacity equivalent to the number of servers. The steady state solution for the M/D/n/n queue is well known as derived by A. K. Erlang. This solution was later shown to be valid for M/G/n/n queues [6].

In this paper, a framework for modelling G/D/n/n queues is presented and a close approximation of and bounds on the blocking probability are derived. It is also shown how the framework is applicable to general service times distributions, that is, G/G/n/n queues. The approximation is shown to comply with simulation for Markovian and Degenerate arrival distributions. In the Markovian case, this entails that it also fits well with well-known exact solutions for M/G/n/n queues.

The presented model and investigation of the G/D/n/n queue is applicable more generally, beyond the motivating scenarios with MTC.

2 System model

We will develop a general analytical model for queues with deterministic service times, but we will treat the concrete problem of reservations of demodulation paths in LoRaWAN gateways as mentioned in the introduction. SX1301 is a chipset meant for usage in LoRaWAN gateways. This chipset is capable of demodulating up to 8 frames in parallel [7].

2.1 Arrival process

Let the number of messages transmitted within a fixed time T be denoted k. The probability of at-least k_0 transmissions within a time step of τ is

$$B_0(\lambda,\tau) = \Pr(k \geq 0 | \sum_{x\in\{\}} t_x \leq \tau, \lambda) := 1 \ , \tag{D.1}$$

$$B_k(\lambda,\tau) = \Pr(k \geq k_0 | \sum_{x=1}^{k_0} t_x \leq \tau, \lambda) = \iint_D (f_{1,2,\ldots,p})\, dD \ ,$$

$$D \in \{t_1, \ldots, t_x | \sum_{x=1}^{k_0} t_x \leq \tau\} \ .$$

Then we can find the probability of transmitting exactly k messages in a period τ by

$$A_k(\lambda,\tau) = \Pr(k = p | \sum_{x=1}^{p} t_x \leq \tau, \lambda) \ , \tag{D.2}$$

$$= \Pr(k \geq p | \sum_{x=1}^{p} t_x \leq \tau, \lambda) - \Pr(k \geq p+1 | \sum_{x=1}^{p+1} t_x \leq \tau, \lambda) \ ,$$

$$= B_k(\lambda,\tau) - B_{k+1}(\lambda,\tau) \ , \quad \text{for} \quad 0 \leq k \leq \infty \ .$$

The set of received messages is a subset of the set of transmitted messages due to outage caused by for example poor channel conditions, noise or interference. Let p_o be the outage probability for a transmission not to be received. Then the probability for the number of received messages can be found by transforming the probability of the number of transmissions as

$$A_k^{\diamond}(\lambda,\tau) = \sum_{x=k}^{\infty} (A_k(\lambda,\tau)(1-p_o)^k p_o^{x-k} \binom{k}{x}) \ .$$

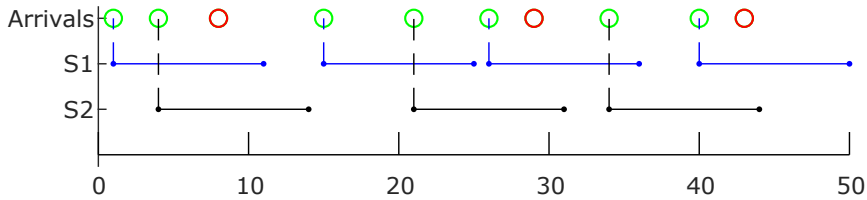


Fig. D.1: Example of arrivals, service times and blocking in a G/D/2/2 queue. Messages occupy servers for a fixed service time, τ. Messages who arrive to find all servers occupied are blocked.

2.2 Queue behaviour

We denote the number of transmissions being demodulated at t_{i-1} by K_{i-1} and the number transmissions taking up demodulation paths after $t_{i-1}+\tau$ by K_i. Denote new arrivals in the queue by K_i^{A}, demodulated transmissions by K_i^{D} and the number of blocked transmissions by K_i^{B}. transmissions are blocked when they arrive to find all demodulation paths unavailable as depicted in Fig D.1. Then we have

$$K_i = K_{i-1} + K_i^{\mathrm{A}} - K_i^{\mathrm{D}} - K_i^{\mathrm{B}} \ , \qquad 0 \le K_i \le n \ . \tag{D.3}$$

$$\begin{aligned} K_i^{\mathrm{B}} &= \max(K_{i-1} + Z_i - n, 0) \ , \\ \text{where} \quad Z_i &= K_i^{\mathrm{A}} - K_i^{\mathrm{D}} \ , \\ \text{so} \quad f_{Z_i} &= f_{K_i^{\mathrm{A}}} * \hat{f}_{K_i^{\mathrm{D}}} \ , \\ \text{where} \quad \hat{f}_{K_i^{\mathrm{D}}}[k] &= f_{K_i^{\mathrm{D}}}[-k] \ , \\ \text{and} \quad f_{K_i^{\mathrm{A}}}[k] &= A_k^{\diamond} \ . \end{aligned} \tag{D.4}$$

All messages that arrived within the prior period and weren't blocked are served, so

$$K_i^{\mathrm{D}} = \min(K_{i-1}^{A}, n) \tag{D.5}$$

3 Analysis

3.1 Bounds on the blocking probability

The blocking probability is defined as

$$P_b = \frac{\mathrm{E}[K_i^{\mathrm{B}}]}{\mathrm{E}[K_i]} = \frac{\mathrm{E}[K_i^{\mathrm{B}}]}{\mathrm{E}[K_i^{\mathrm{B}}] + n} \ , \tag{D.6}$$

where the substitution $\mathrm{E}[K_i] = \mathrm{E}[K_i^{\mathrm{B}}] + n$ in the denominator in (D.6) is valid, because the blocking probability is zero until the amount of messages in queue is larger than the total number of demodulation paths n. This also means that

$$f_{K_i^{\mathrm{D}}}[k] = \begin{cases} 1\,, & \text{for } k = n\,, \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{D.7}$$

Assume that there's no spill-over between messages in the observed periods $i-1$ and i, then $K_{i-1} = 0$ and we obtain a lower bound on the blocking probability.

$$\begin{aligned} P_{b_{\text{lower}}} &= \frac{\sum_{k=1}^{\infty} f_{K_i^{\mathrm{B}}|K_i=n}[k] \cdot k}{\sum_{k=1}^{\infty} f_{K_i^{\mathrm{B}}|K_i=n}[k] \cdot k + n}\,, \\ &= \frac{\sum_{k=n+1}^{\infty} (A_k(\lambda,\tau) \cdot (k-n))}{\sum_{k=n+1}^{\infty} (A_k(\lambda,\tau) \cdot (k-n)) + n}\,. \end{aligned} \tag{D.8}$$

In the same manner, we can assume that there's complete spill-over between messages in period $i-1$ and i, then $K_{i-1} = n$ and we obtain an upper bound on the blocking probability.

$$\begin{aligned} P_{b_{\text{upper}}} &= \frac{\sum_{k=1}^{\infty} f_{K_i^{\mathrm{B}}|K_i=0}[k] \cdot k}{\sum_{k=1}^{\infty} f_{K_i^{\mathrm{B}}|K_i=0}[k] \cdot k + n}\,, \\ &= \frac{\sum_{k=n}^{\infty} (A_k(\lambda,\tau) \cdot k)}{\sum_{k=n}^{\infty} (A_k(\lambda,\tau) \cdot k) + n}\,. \end{aligned} \tag{D.9}$$

3.2 Blocking probability

To describe the exact blocking probability we need to describe the spill-over between observation period $i-1$ and i, K_{i-1}. We let $X_i = K_{i-1} + Z_i$ so that

$$f_{X_i}[x_i] = \sum_{k_{i-1}=0}^{n} f_{K_{i-1},Z_i}[k_{i-1}, x_i - k_{i-1}]\,, \tag{D.10}$$

$$\begin{aligned} f_{K_{i-1},Z_i}[k_{i-1}, z_i] = \\ f_{Z_i}[Z_i = z | K_{i-1} = k_{i-1}] \cdot f_{K_{i-1}}[K_{i-1} = k_{i-1}]\,. \end{aligned} \tag{D.11}$$

Then we can describe the probability of blocking k transmissions as

$$f_{K_i^{\mathrm{B}}}[k] = \begin{cases} \sum_{x=-\infty}^{n} f_{X_i}[x] \ , & \text{for } k = 0 \ , \\ f_{X_i}[k+n] \ , & \text{for } k \geq 1 \ , \\ 0 \ , & \text{otherwise} \ . \end{cases} \tag{D.12}$$

Then we can use a prior for $f'_{K_{i-1}}$ to approximate $f'_{K_i^{\mathrm{D}}}$, f'_{Z_i} and $f_{X_i} \approx \dfrac{f_{X_i}|f'_{K_{i-1}}}{\sum(f_{X_i}|f'_{K_{i-1}})}$ and then approximate the blocking probability, P_b, using (D.6).

In case a prior is not evident for an arrival process G, then we may use

$$f_{K_{i-1}}[k] = \begin{cases} \sum_{x=-\infty}^{0} f_{K_i^{\mathrm{A}}}[X_i = x] \ , & \text{for } k = 0 \ , \\ f_{X_i}[K_i^{\mathrm{A}} = k] \ , & \text{for } 0 < k < n \ , \\ \sum_{x=n}^{\infty} f_{K_i^{\mathrm{A}}}[X_i = x] \ , & \text{for } k = n \ , \\ 0 \ , & \text{otherwise} \ . \end{cases} \tag{D.13}$$

3.3 Server state probability and server utilization

The server utilization can be found by considering the timing within the queue. Consider the case of $n = 1$, then the time spent without messages in queue can be described as the time between completion of service of one message till the arrival of the next. Hence:

$$T_0 = \sum_{k=0}^{\infty} C_k \quad \text{and} \quad T_1 = \tau \ , \tag{D.14}$$

where C_k is the mean time spent without a message in queue if the k'th message is the first one received after the demodulation of another message finishes,

$$C_k = \int_{t=\tau}^{\infty} \Pr(\sum_{x=1}^{k+1} t_x = t | \sum_{x=1}^{k} t_x \leq \tau) \cdot (t - \tau) \, dt \ , \quad \text{for} \quad k \geq 0 \ . \tag{D.15}$$

The complexity of describing the timing in this way increases greatly as n increases. The state ratio for state y is given by

$$q_y = \frac{T_y}{\sum T_y} \ . \tag{D.16}$$

Based on the state ratios we can compute the average number of messages being served as (D.17) and the server utilization as (D.18).

$$L = \sum_{y=0}^{n} q_y \cdot y \ . \tag{D.17}$$

$$\eta = \frac{L}{n} \ . \tag{D.18}$$

$$\zeta = \eta(1 - P_b) \ , \tag{D.19}$$

where ζ is the non-blocking server utilization, which is the probability that the demodulation path is being used and message demodulation is not being blocked.

3.4 Jobs with non-homogeneous service times

We divide messages into classes C_1 through C_m corresponding to the different spreading factors in LoRaWAN. The service time of class x is given as $\tau_x \in \boldsymbol{\tau} = \{\tau_1, \tau_2, ..., \tau_m\}$ for corresponding mean arrival rates of $\lambda_x \in \boldsymbol{\lambda} = \{\lambda_1, \lambda_2, ..., \lambda_m\}$.

Denoting the number of messages from class y by k_y, the probability of k_y messages arriving within the service period, τ_y, is $A_k(\lambda_y, \tau_y)$ and the arrival distribution $f_{K_i^{\mathrm{A}}}$, we have

$$f_{K_i^{\mathrm{A}}}(\boldsymbol{\lambda}, \boldsymbol{\tau}) = \sum_{y=1}^{m} f_{K_i^{\mathrm{A}}}(\lambda_y, \tau_y) \frac{\lambda_y}{\sum_{x=1}^{m} \lambda_x} \ . \tag{D.20}$$

Any service time distribution can be represented by binning with infinitesimally small bins, so that we obtain two infinite sets for $\boldsymbol{\lambda}$ and $\boldsymbol{\tau}$, where $\lambda_x = p_x \cdot \lambda$ and $p_x = f_{\mathrm{ServiceDist}}(\tau_x)$. Practically, the service time distribution can be accurately approximated by binning with an appropriately small finite bin size. In this way we may be able to represent any G/G/n/n queue in the described framework by binning the service time distribution.

4 Degenerate and Markovian arrival processes

In this section the transmission count probability, A_k, is derived for degenerate and Markovian arrival processes by solving (D.2). The mean arrival-rate, λ', and priors, $f'_{K_{i-1}}$, are also discussed for each arrival process.

4.1 Degenerate inter-arrival distribution

Let transmissions occur with inter-arrival times t_x that are distributed according to a degenerate distribution; $f_t(t) = p_x$ for $t = t_x$ where $t_x \in \{t_1, t_2, ..., t_y\}$ for a corre-

sponding set $p_x \in \{p_1, p_2, ..., p_y\}$ where $\sum_{x=1}^{y} p_x = 1$. Solving (D.1) and (D.2) for the D/D/n/n queue we obtain

$$
\begin{aligned}
B_0(\lambda,\tau) &= \sum_D p_x \ , \quad \text{for} \quad D = \{x | t_x \geq \tau\} \ , \\
B_1(\lambda,\tau) &= \sum_D p_x \ , \quad \text{for} \quad D = \{x | t_x \leq \tau\} \ , \\
B_2(\lambda,\tau) &= \sum_D \prod_{i=1}^{2} p_{x_i} \ , \ \text{for } D = \{x_1, x_2 | \sum_{y=1}^{2} t_{x_y} \leq \tau\} \ , \\
B_k(\lambda,\tau) &= \sum_D \prod_{i=1}^{k} p_{x_i} \ , \ \text{for } D = \{x_1, x_2, ..., x_k | \sum_{y=1}^{k} t_{x_y} \leq \tau\} \ , \\
A_k(\lambda,\tau) &= B_k(\lambda,\tau) - B_{k+1}(\lambda,\tau) \ .
\end{aligned} \tag{D.21}
$$

The prior of (D.13) gives good approximations for $n > 1$. When n=1, it is clear that $K_{i-1} = 0$ holds, so our prior should be $f_{K_{i-1}}[k] = 1$ for $k = 0$ where $f_{K_{i-1}}[k] = 0$. for $k \neq 0$ The mean arrival-rate is $\lambda' = \dfrac{1}{\sum_{x=1}^{y} \dfrac{p_x}{t_x}}$. C_k can be found to be

$$
\begin{aligned}
C_k(\lambda,\tau) = \sum_D (\prod_{i=1}^{k+1} p_{x_i} \cdot (\sum_{i=1}^{k+1} t_{x_i} - \tau)) \ , \\
\text{for } D = \{x_1, x_2, ..., x_{k+1} | \sum_{y=1}^{k} t_{x_y} \leq \tau, \sum_{y=1}^{k+1} t_{x_y} > \tau\} \ ,
\end{aligned} \tag{D.22}
$$

4.2 Exponential inter-arrival distribution

Let transmissions occur with inter-arrival times t_x that are distributed according to an exponential distribution; $f_t(t) = \lambda \exp(-\lambda t)$ for $t \geq 0$. Solving (D.2) for this arrival process we obtain

$$
\begin{aligned}
A_0(\lambda,\tau) &= \exp(-\lambda\tau) \ , \\
A_1(\lambda,\tau) &= \lambda\tau \cdot \exp(-\lambda\tau) \ , \\
A_2(\lambda,\tau) &= \frac{(\lambda\tau)^2}{2} \cdot \exp(-\lambda\tau) \ , \\
A_3(\lambda,\tau) &= \frac{(\lambda\tau)^3}{6} \cdot \exp(-\lambda\tau) \ , \\
A_k(\lambda,\tau) &= \frac{(\lambda\tau)^k}{k!} \cdot \exp(-\lambda\tau) \ .
\end{aligned} \tag{D.23}
$$

ID	n	t_1	t_2	t_3	p_1	p_2	p_3	p_o
1	1	0.3	0.6	1.5	1/3	1/3	1/3	0
2	2	0.3	0.6	1.5	1/3	1/3	1/3	0
3	1	0.3	0.6	1.5	1/3	1/3	1/3	0.5
4	2	0.3	0.6	1.5	1/3	1/3	1/3	0.5

Table D.1: Various degenerate arrival process configurations.

The mean arrival-rate is λ. We shall use the prior in (D.13) when approximating P_b. C_k can be found to be

$$C_k(\lambda, \tau) = \frac{\exp(-\lambda\tau)\lambda^{k-1}\tau^k}{k!} \,. \tag{D.24}$$

5 Results

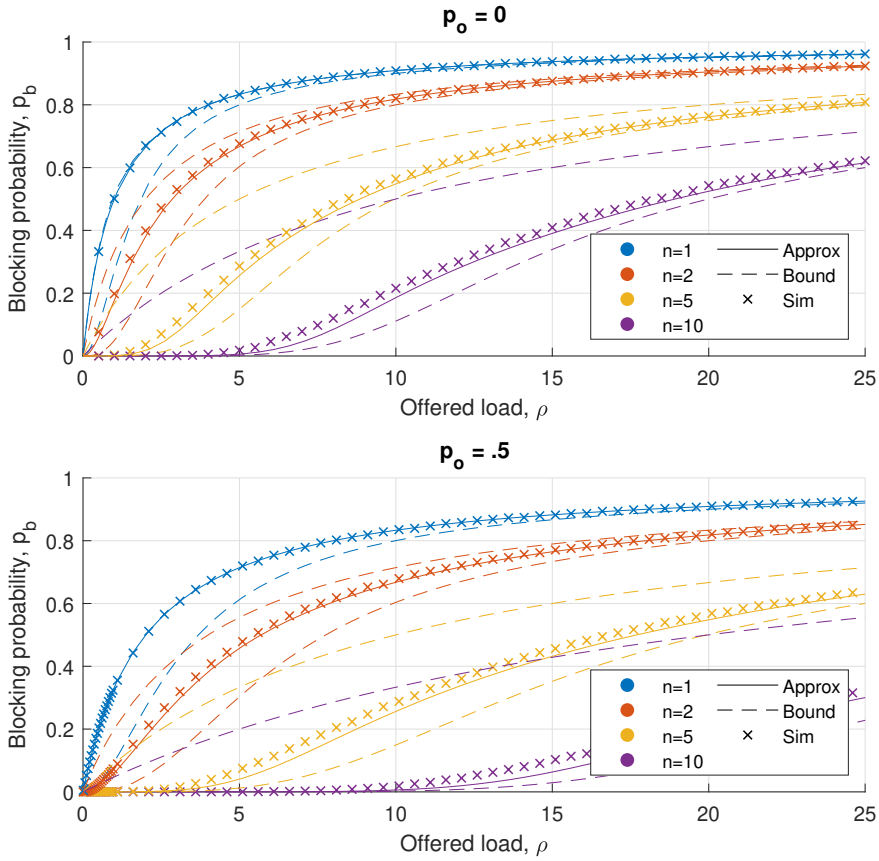
In this section we present results for the accuracy of this framework for modelling M/D/n/n and D/D/n/n queues for a fixed service time and a set of heterogeneous service times. Then we discuss the impact of the results on the example case of blockage in the demodulation paths of a LoRaWAN gateway.

The approximated blocking probability and bounds for the M/D/n/n queue can be found in Fig. D.2a. The approximation is close to the exact value, i.e. Erlang-B result. The server efficiency and non-blocking server efficiency are plotted in Fig. D.3a. Here the result is exact owing to the timing analysis in Sec. 3.3 for $n = 1$.

The arrival count as a function of the service time in a D/D/n/n queue changes as a nontrivial step-wise function of the service time τ and inter-arrival rate as depicted in Fig. D.3b. We observe that the index x of the smallest neglible A_x grows with the offered traffic load in Erlang. This also applies to the M/D/n/n queue, but in that case the count probability is a smooth function. The blocking probability exhibits the same behaviour as depicted in Fig. D.4a. The approximated blocking probability, here is also very close to simulated values. The Server efficiency and non-blocking server efficiency can be found in Fig. D.5a. Notice that the mean arrival rate in the analysis of the D/D/n/n queue is fixed at $\lambda' = 1/(0.3 \cdot 1/3 + 0.6 \cdot 1/3 + 1.5 \cdot 1/3\ ) = 1.25$.

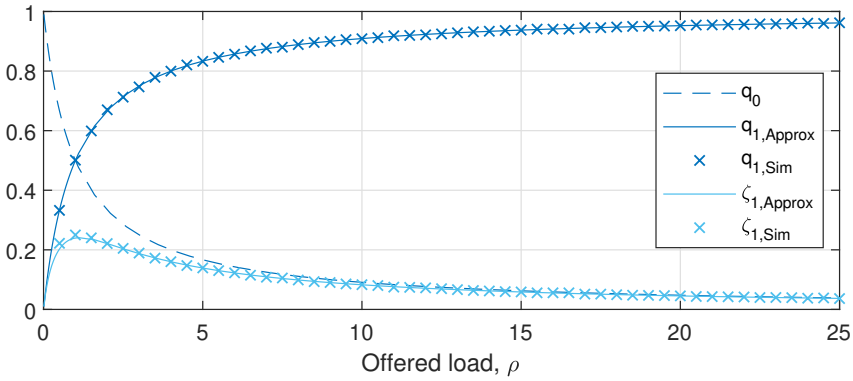
In Fig. D.5b the blocking probabilities for the M/D/n/n queue and D/D/n/n queue are plotted for a service process that is defined by $p_x \in \boldsymbol{p} = \{\frac{1}{2}, \frac{1}{2}\}$ and $\tau_x \in \boldsymbol{\tau} = \{\frac{2}{3}, \frac{4}{3}\}$. The results are a close approximation, indicating the (D.20) holds.

The results of our investigation are straightforward to interpret as trade-offs in our motivating example of the blocking probability of demodulation paths in a LoRa receiver; Clearly a larger number of demodulation paths n yields a lower blocking probability,

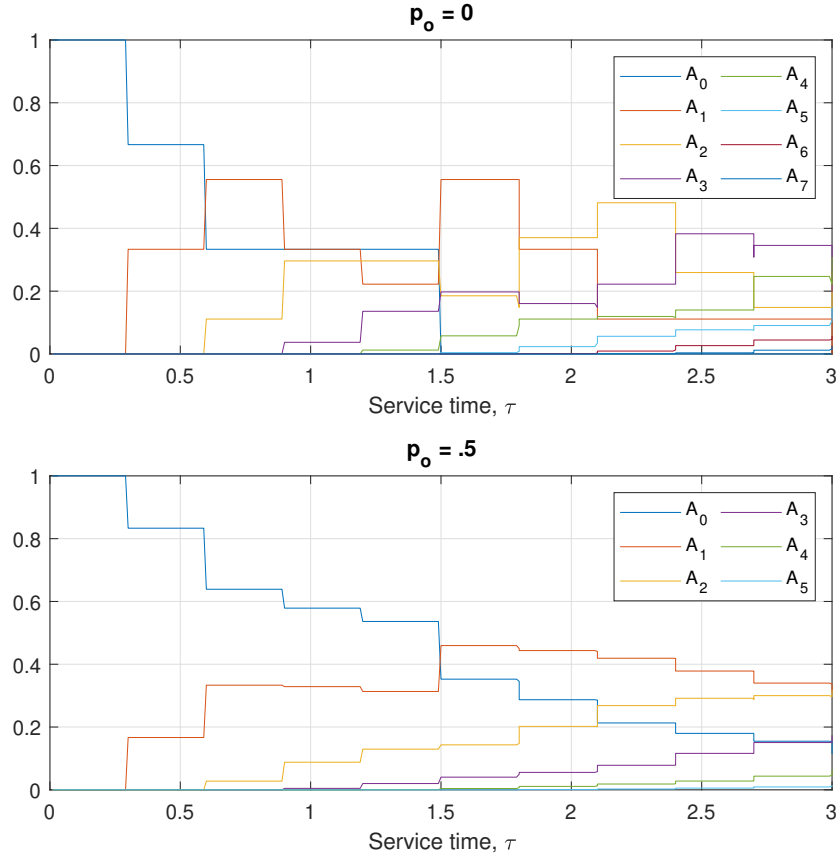


(a) The blocking probabilities for the M/D/n/n queue for expected offered load ρ and intermittent arrival processes for $p_o = 0$ and $p_o = 0.5$. Notice that bounds hold and the approximation is close to the exact solution.

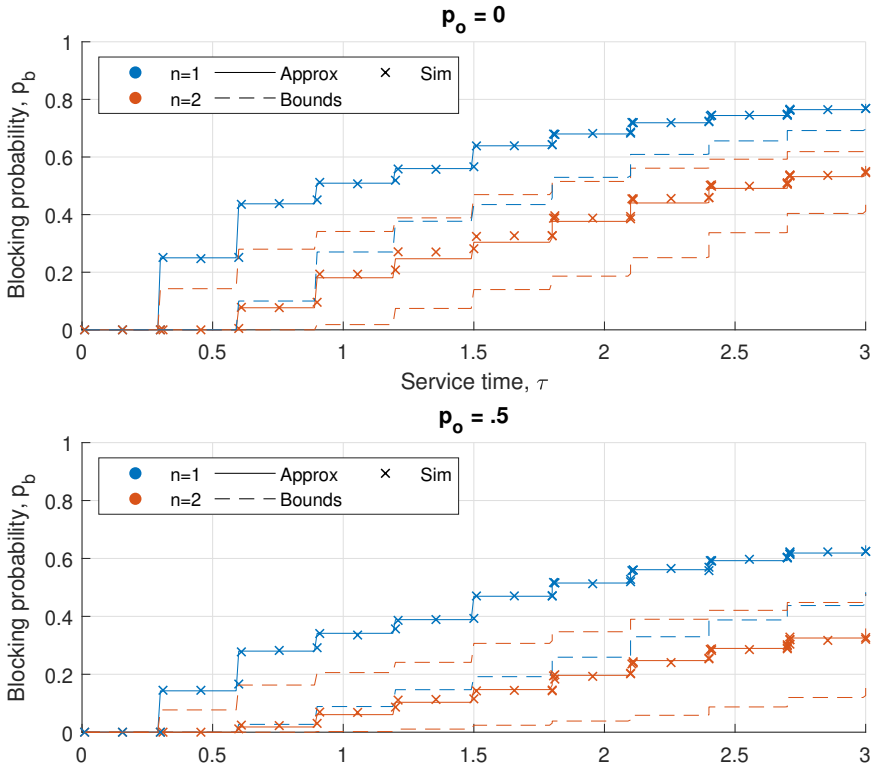
however the cost of increasing n should yield an equivalent gain in performance. In general the distribution of the arrival-process is of course of dire importance. We observe a higher blocking probability $P_b = .25$ in the M/D/2/2 queue than $P_b = .068$ in the D/D/2/2 queue (configured as ID 2 in D.1) for $\tau = 1$ and $\lambda' = .8$. It is also evident that given Degenerate arrival times tuning the service time relatively little can yield large gains.



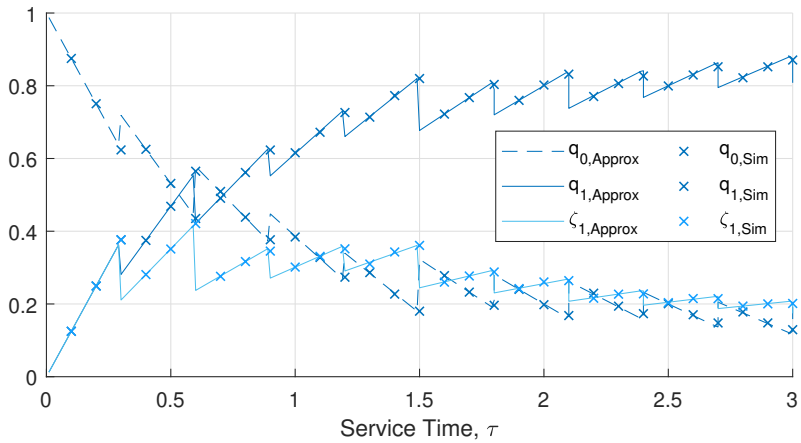
(a) State probabilities, q_0 and q_1, server efficiency, $\eta = q_1$, and non-blocking server efficiency, ζ, for the M/D/1/1 queue.



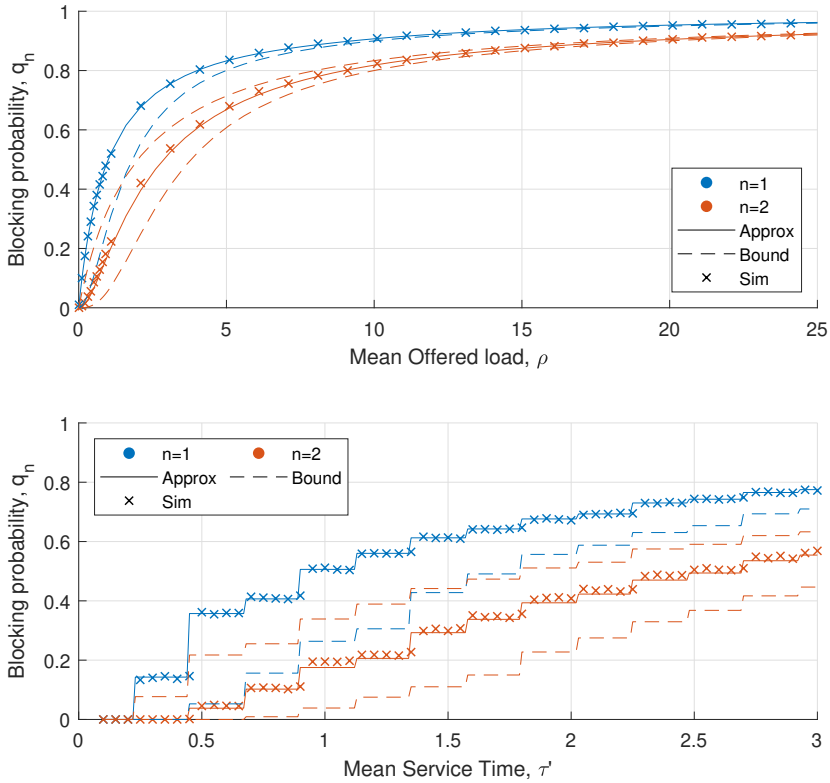
(b) The probabilities A_x of x arrivals within the service time τ in D/D/n/n queues for the configurations of the degenerate arrival process outlined in Tab. D.1. A_x does not depend on the number of servers n so the results are equivalent for configurations ID 1 & 3 and ID 2 & 4, respectively.



(a) The blocking probability in D/D/n/n queues as a step-wise function of the service time τ for the four configurations of the degenerate arrival process outlined in Tab. D.1.



(a) State probabilities, q_0 and q_1, server efficiency, $\eta = q_1$, and non-blocking server efficiency, ζ, for the D/D/1/1 queue for ID 1 in Tab. D.1.



(b) P_b for Markovian (above) and degenerate (below) arrival processes and a service process that is defined by $p_x \in \boldsymbol{p} = \{\frac{1}{2}, \frac{1}{2}\}$ and $\tau_x \in \boldsymbol{\tau} = \{\frac{2}{3}, \frac{4}{3}\}$. The degenerate arrival processes in this example is ID 1&ID 2 where $\boldsymbol{\tau}$ is scaled to achieve the mean service rate.

6 Conclusion

In this paper, a framework for modelling G/D/n/n queues was presented and we looked at the case of blocking demodulation paths in LoRaWAN receivers. The proposed framework assesses the arrival count in service periods to model blocking probabilities in G/D/n/n quite accurately. Bounds valid for the blocking probability in G/D/n/n queues were also presented. In essence, we showed how blocking probability depends on the arrival process, which is therefore essential for describing the exact blocking probability for e.g. demodulation in LoRaWAN receivers. In general, increasing n, decreases the blocking probability as one would expect intuitively. Exponential inter-arrival times are often assumed for the arrival process of communication networks, but for example in the case of the LoRaWAN receiver, this transmission process will be filtered by capture effect, yielding a non-Markovian distribution at the demodulation paths.

The framework depends on finding counting functions, which may be relatively easy to find by induction given the tools available today for integrating symbolic expressions. Moreover, the framework is directly applicable when the distribution of the inter-arrival times is given numerically as 'binned' or Degenerate approximation of the arrival process, but it is not available on an explicit analytical form. The methodology was shown to yield close approximates of the well known results for the M/G/n/n queue along with close approximates for the D/D/n/n queue. The framework was extended to G/G/n/n-queues, but verification was limited to a small set of service times due to simulation complexity.

References

[1] T. Hoßfeld, F. Metzger, and P. E. Heegaard, "Traffic modeling for aggregated periodic iot data," in *2018 21st Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*, 2018, pp. 1–8.

[2] A. Rahmadhani and F. Kuipers, "When lorawan frames collide," *Proc. of the 12th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (ACM WiNTECH 2018)*, 2018.

[3] R. B. Sørensen, N. Razmi, J. J. Nielsen, and P. Popovski, "Analysis of lorawan uplink with multiple demodulating paths and capture effect," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, May 2019, pp. 1–6.

[4] J. Kreuzinger, A. Schulz, M. Pfeffer, T. Ungerer, U. Brinkschulte, and C. Krakowski, "Real-time scheduling on multithreaded processors," in *Proceedings Seventh International Conference on Real-Time Computing Systems and Applications*, 2000, pp. 155–159.

[5] D. G. Kendall, "Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain," *Annals of Mathematical Statistics*, vol. 24, no. 3, pp. 338–354, 09 1953. [Online]. Available: https://doi.org/10.1214/aoms/1177728975

[6] L. Takacs, "On erlang's formula," *The Annals of Mathematical Statistics*, vol. 40, no. 1, pp. 71–78, 1969. [Online]. Available: http://www.jstor.org/stable/2239199

[7] Semtech, *SX1301 Datasheet*, v2.4 ed., June 2017.

Paper E

Machine Learning Methods for Monitoring of Quasi-Periodic Traffic in Massive IoT Networks

René Brandborg Sørensen, Jimmy Jessen Nielsen, Petar Popovski

The paper has been published in the
IEEE Internet of Things Journal, vol. 7, no. 8, pp. 7368–7376, Aug. 2020.

Abstract

One of the central problems in massive Internet of Things (IoT) deployments is the monitoring of the status of a massive number of links. The problem is aggravated by the irregularity of the traffic transmitted over the link, as the traffic intermittency can be disguised as a link failure and vice versa. In this work we present a traffic model for IoT devices running quasi-periodic applications and we present unsupervised, parametric machine learning methods for online monitoring of the network performance of individual devices in IoT deployments with quasi-periodic reporting, such as smart-metering, environmental monitoring and agricultural monitoring. Two clustering methods are based on the Lomb-Scargle periodogram, an approach developed by astronomers for estimating the spectral density of unevenly sampled time series. We present probabilistic performance results for each of the proposed methods based on simulated data and compare the performance to a naïve network monitoring approach. The results show that the proposed methods are more reliable at detecting both hard and soft faults than the naïve-approach, especially when the network outage is high. Furthermore, we test the methods on real-world data from a smart metering deployment. The methods, in particular the clustering method, are shown to be applicable and useful in a real-world scenario.

1 Introduction

Iot deployments can provide a large variety of services capable of cyber-physical interactions through sensors, actuators and data analysis by utilizing fog or cloud computing. Such deployments can be cyber-physical systems, consisting of devices that exchange messages with servers through networks, as depicted in Fig. E.1, which also shows the common IoT architecture [1]. In such deployments the traffic generated by sensors is intermittently filtered by the network before being received by the IoT server. This intermittency is caused random network effects, such as medium access delays, outage in the network and queuing of transmissions.

During the past decade, techniques and standards have been developed to provide adequate networking features and improve the Quality of Service (QoS) for the vast number of IoT use cases. An overview of the architecture of IoT services and enabling technologies and protocols is given in [1]. Specifically, in the context of wireless communications, the term IoT usually refers to *massive machine-type communication* (mMTC), one of the three connectivity types in 5G [2]. Here, a number of pre-5G IoT systems for massive connectivity have been developed and are currently being deployed. These include the low power wide area networks (LPWANs): SigFox, LoRaWAN, NB-IoT and LTE-M [3–5].

A central problem of wireless IoT connectivity is monitoring and status detection for a massive number of connected devices, which provides insights into the status of

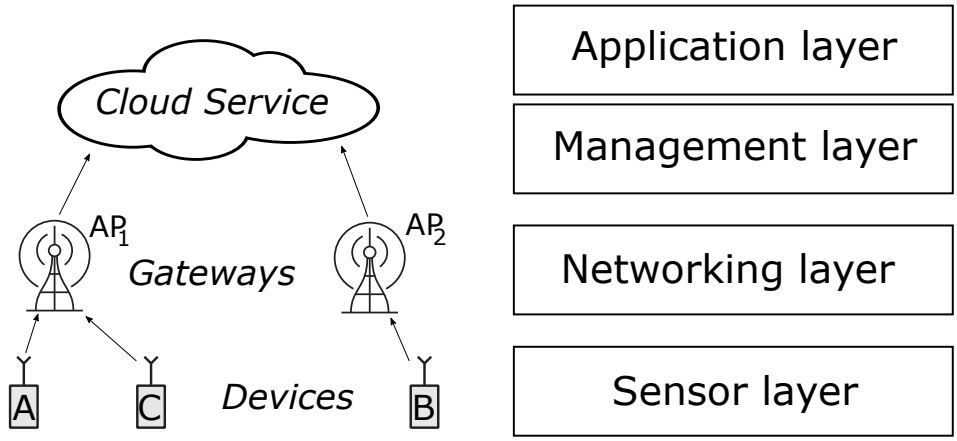


Fig. E.1: ©2020 IEEE: IoT deployment topology and architecture. A gateway could be, for example, a cellular towers, a LPWAN access point, or a satellite. The network topology between devices and gateways may be a mesh or a star-topology. The IoT deployment is connected to a virtual IoT server in the cloud, which manages the network and makes application features available to end users. The different tiers of the network are directly sensing different information about the network. Information regarding link quality of connected devices is available to the APs, but only to the management tier if this information is relayed.

the links and devices and can potentially lead to corrective actions [1, 6, 7]. Network monitoring for wireless sensor networks (WSN) has been researched for decades. A comprehensive survey of network monitoring in wireless sensor networks (WSNs) can be found in [6] and [7]. Arguments for the importance of network monitoring is given by these surveys; network monitoring can enhance data reliability, bandwidth utilization, and the lifetime of the WSN due to the opportunity to identify faulty devices and hence better utilize constrained resources. Multiple monitoring methods can be combined through frameworks, such as fuzzy logic [8] to optimize decision making for fault tolerance. In general, the fault detection methods for WSNs assume a PAN mesh topology such as 6LoWPAN or ZigBee. For example, PAD, a passive monitoring method relying on inference based on routing changes is presented in [9]. Nevertheless, LPWANs are one-hop star-topology networks and additionally, PAD introduces a few bytes of overhead to transmissions, which is a major drawback for energy-constrained devices and networks supporting massive numbers of devices. Notably, the fault detection methods of [10] and [11] rely on statistical inference based on the timing of incoming traffic to detect faults in the network. These methods may readily be modified for and applied to LPWANs since they do not rely on topology changes in a mesh network, but they have the drawback of requiring fault-less training data in order to recognise healthy behavior and the detected errors are only quantifiable at a low resolution, ie. "no errors, some errors or many errors".

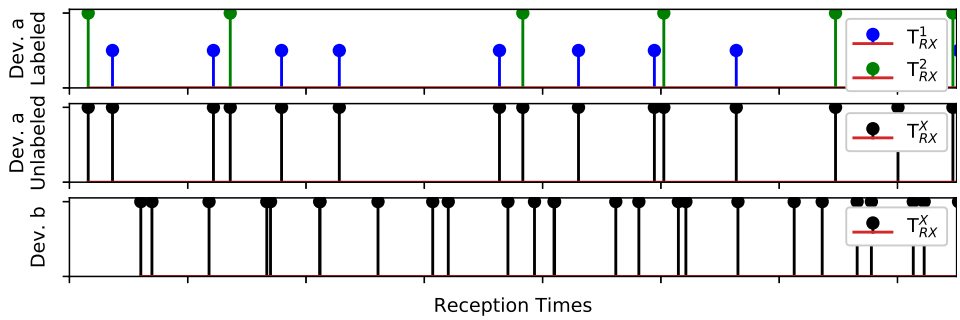


Fig. E.2: ©2020 IEEE: Devices generate traffic depending on their applications. Here the traffic patterns are depicted for three types of devices; Dev. a) runs two applications that generate quasi-periodic traffic. The received traffic streams are shown for both labeled and unlabeled traffic. Dev. b) runs three quasi-periodic applications, but does not label the traffic.

In this paper, we make no assumption about the topology of the network and instead we assume a quasi-periodic[1] traffic model as that depicted in Fig. E.2 for device A and B, which is common for IoT applications. This approach allows us to evaluate the state and link performance of individual devices passively, which is important for identifying poor performers and malfunctioning devices in massive IoT deployments. In our model devices can run 'thin' clients (a single application) or 'thick' clients (multiple applications). We have observed the latter behaviour in a data-set from a LoRaWAN deployment, where mains-powered sensors and actuators were used to control and manage street lights in rural towns. We propose parametric machine learning methods for high resolution, centralised and passive fault detection in arbitrary IoT deployments. The methods use temporal correlations in observed traffic to parameterize quasi-periodic applications. The modelled applications are used to infer the state and the QoS of individual devices. Interestingly, the methodological basis for this work has been drawn from research in astronomy, where unevenly sampled time series are common. Astronomers have developed techniques for analyzing such series including phase-folding [12] and a variant of the classical Fourier periodogram that is generalised for uneven time-series [13–16].

The paper is structured as follows: We introduce our traffic model, the traffic metadata that can be expected to be available for analysis, and targeted network KPIs in Sec. 2. In Sec. 3.1 we analyze the sub-problem of parametric regression for traffic that is labeled by its parent application. The classification problem of labeling traffic when the traffic parameters are known, but the parent applications are unknown, is examined in Sec. 3.2. In Sec. 4 we treat the clustering problem that arises when no a priori information is given. Performance results for the algorithms presented throughout can

[1] Quasi-periodic applications generate transmissions at a constant, or near constant, inter-arrival time. See Sec. 2.2 for more details.

be found in Sec. 5 and the results of using the fault-detection algorithms on a real-world smart metering deployment can be found in Sec. 6. Sec. 7 contains concluding remarks. Table E.2 on page 109 provides a list of symbols and mathematical notations.

2 System Model and Key Performance Indicators

Consider an IoT deployment like that of Fig. E.1 & Fig. E.2 where sporadically transmitting devices are connected to a server by an arbitrary networking technology, for example an LPWAN, or a mesh network. Here, a virtual server running in the cloud acts as a centralised management layer for network monitoring. The wireless network acts as a filter upon transmitted data introducing intermittency in the form of outage and delays in the received data. In this section we generalise this model to any number of gateways and devices running any number of applications. First, we present the meta-data available in a passive monitoring scenario, define a traffic model for quasi-periodic IoT devices, a set of key performance indicators (KPIs) and a naïve method for evaluating the KPIs. Lastly, we examine how meta-data changes the nature of the machine learning problem.

2.1 Available Traffic Meta-Data

In Fig. E.1 the devices A and C transmit data through AP_1 while device B transmits data through AP_2. The data that is available for analysis at the server depends on how much meta-data AP_1 and AP_2 relay to the server in addition to how much meta-data is included in the transmissions. When a network is licensed from a network provider then it is considered to be a public network [17]. A network that is closed to the public, privately owned or purpose-built specifically for an IoT deployment, is considered a private network. The same network can be considered private to its owner and public to licencees. In private networks link level metrics, such as received signal strength indicator, link quality indicator, signal to noise ratio, channel state information, modulation and coding rate, are available whereäs they are not necessarily available in public networks. Some network technologies and protocols have specific meta-data built in to the protocol, for example GPS coordinates in SigFox. Such metrics can not always be expected to be available in public networks.

We define a minimal set of metrics that are available in any type of network, {Network ID, Device ID, Reception Timestamp, Payload size}. Device and Network identifiers are a necessary part of a useful transmission. The transmission size and reception time can likewise be found for any transmission. This ensures that the network monitoring is applicable both for network operators, who licence their networks to IoT deployments, but do not have in-depth knowledge of the IoT deployment or access to transmitted data and IoT vendors and operators, who do not have insight into the networking components or access to link-level meta-data.

In addition to this minimal set of metrics, devices may include the ID of their parent application as meta-data within the transmission. This is the behavior of device A in Fig. E.2 in contrast to device C, which does not label its traffic. The implication of labeling is that we may pose the fault detection problem as a regression problem instead of a clustering problem. We discuss this implication further in Sec. 2.5 after defining our traffic model and KPIs.

2.2 Traffic model

The traffic generated by a device is the composite of the traffic generated by all the applications running on the device. We denote the number of apps running on a device by I_{apps}. Traffic of one app may influence the jitter of traffic of another app due to queuing of transmissions in the device.

$$T^x_{\text{RX}} = T^1_{\text{RX}} \cup T^2_{\text{RX}} \cup ... \cup T^{(i)}_{\text{RX}} \cup ... \cup T^{I_{\text{apps}}}_{\text{RX}} \tag{E.1}$$

Where $T^i_{\text{RX}} = \{T^i_1, T^i_2, ..., T^i_m, ...T^i_{M-1}, T^i_M\}$ denotes the received transmissions from application i.
We define Quasi-periodic applications:

These are applications, which send periodic reports, but where the received traffic is intermittent due to queuing and filtering by the network. Common IoT use cases such as gas-, water- and electric smart metering, smart agriculture and smart environment [18, 19], are considered by 3GPP to be quasi-periodic [18].

The ith quasi-periodic app generates transmissions at approximately constant intervals, such that the reception times of transmissions from app i can be described by:

$$T^i_{m^i} = \beta^i + \alpha^i \cdot (m^i + o^i_{m^i}) + J^i_{m^i} \tag{E.2}$$

where β^i is a time offset, α^i is the inter-arrival time of app i, J^i_m is a random delay introduced by the network (jitter), m^i is the index of the received packets while $o^i_{m^i}$ is the cumulative number of transmissions that were not received until observation m^i. Let n^i be the index of the transmissions such that $n^i = m^i + o^i_{m^i}$. Then we know that $T^i_{n-1} \cong T^i_n \cong T^i_{n+1} \pmod{\alpha^i}$ for the quasi periodic app i.

2.3 Key Performance Indicators

We wish to monitor the link quality and status of individual devices in IoT networks. Based on the available meta-data we choose to monitor network outage and online/offline status, which are common KPIs in wireless network performance modelling. In this subsection we define each of these KPIs.

Offline detection

An app or a device is considered to be offline if it stops generating transmissions. An intuitive classifier for offline status is detecting whether k consecutive expected transmissions have been missed. This can be done at the application level to classify applications as offline or at the device level, to classify an entire device as offline. Offline applications are not expected to generate transmissions and so they do not count towards the calculated outage. We define a classifier for offline entities, C_{Off}.

$$C_{\text{Off}} : o^i(t) \geq k \tag{E.3}$$

where $o^i(t) = \left\lfloor \frac{t - T^i_{m^i}}{\alpha^i} \right\rfloor$ computes the expected number of transmissions at time t since the last reception at $T^i_{m^i}$ if the quasi-periodic app i was online.

Outage probability

We define the outage probability as the ratio of the number of packets lost to the number of transmitted packets over a window τ_w at time t.

$$p^*_{\text{o}} = \sum_{i=1}^{I_{\text{apps}}} o_{m^i(t,\tau_w)} \Bigg/ \sum_{i=1}^{I_{\text{apps}}} n^i(t, \tau_w) \tag{E.4}$$

where the number of transmitted and received packets for app i from time $t - \tau_w$ to t are denoted by $n^i(t, \tau_w)$ and $m^i(t, \tau_w)$, respectively. As $n^i(t)$ approaches ∞ the observed outage, p^*_{o}, approaches the network outage probability, p_{o}, if p_{o} is stationary over observation period.

$o_{m^i(t)}$ and $n^i(t)$ are latent variables from the IoT server's perspective. Then the goal of fault detection can be posed as the problem of estimating these latent variables correctly.

2.4 Naïve monitoring method

Here, we introduce a naïve monitoring method, which we will compare other methods to. The approach is straightforward; Denote the number of transmissions received within τ_w at time t, $m_{\text{obs}}(t, \tau_w)$, and let $m_{\max}$ denote the maximum value of $m_{\text{obs}}(t, \tau_w)$ for $t < t_{now}$. Then we have (E.5).

$$p^*_{\text{o}}(t, \tau_w) = \frac{m_{\max} - m_{\text{obs}}(t, \tau_w)}{m_{\max}} \tag{E.5}$$

Devices where $p^*_{\text{o}}(t, \tau_w) > \epsilon$ are classified as offline.

Determining problem nature and applicable methodology		*Known applications and traffic parameters*	
		Yes	*No*
Labelled traffic	*Yes*	KPI calculation	Regression
	No	Classification	Clustering

Table E.1: ©2020 IEEE: Diagram of the classes of the fault detection problem.

2.5 A priori knowledge and labeled traffic

We may have a priori knowledge about online devices and applications, and we may receive meta-data identifying the app source of traffic. This changes the nature of the fault monitoring problem as depicted in Table E.1. In the case where we know the parameters of the traffic model and transmissions are labeled (or all clients are 'thin') it is straightforward to calculate the KPIs. In case the traffic model is unknown, but traffic is labeled we must perform a regression of the traffic parameters in order to calculate the KPIs, which is treated in Sec. 3.1. In Sec. 3.2 we examine the case that traffic parameters are known, but the traffic is unlabeled (and from a 'thick' client) we must classify received packets as belonging to one app or another to evaluate the KPIs. Finally, we tackle the problem of clustering when the traffic parameters are unknown and the traffic is unlabelled in Sec. 4.

3 Regression and Classification

3.1 Regression

In case transmissions are labeled by their parent application, the composite stream of received packets from all applications can easily be sorted by application such that the stream of any periodic application is on the form of (E.2). We wish to learn α^i and $o^i_{m^i}$ given a set of reception times, $T^i_{\mathrm{RX}} = \{T^i_1, T^i_2, ...T^i_m...T^i_{M-1}, T^i_M\}$. Phase-folding methods like phase dispersion minimization [12] could solve this problem in a brute-force manner by estimating the fit of all potential α^i, but the associated computational overhead is undesired for massive online network monitoring.

Instead, we propose Normalised Harmonics Mean (NHM), which is an online method we developed for finding an estimate α^{i*} of α^i in a set T^i_{RX} with a relatively low amount of computational effort.

Consider a MMSE of the distance between a reception time and transmission time, $\min(|T^i_{n^i_{m^i}} - T^i_{m^i}|^2)$ where $n^i_m = m^i + o_{m^i}$. Here o_{m^i} and n_{m^i} are unknown to the

receiver, but we know that the generating function for $T^i_{n_{m^i}}$ is periodic. We may solve the problem by brute-force using least-squares, however this requires finding n_{m^i}, which minimizes the MSE for every arrival m^i for every proposed α^i, which makes this solution computationally expensive and the accuracy depends on the α^i grid chosen for the analysis. This is similar to the brute-force approach of phase dispersion minimization [12].

We propose to use a periodic function, specifically the cosine, to describe the problem. Then we have (E.6) at the transmitter side and (E.7) at the receiver side.

$$\cos\left(\frac{T^i_{n^i_m} - T^i_{n^i_m - 1}}{\alpha^i} \cdot 2\pi\right) = 1 \tag{E.6}$$

$$\cos\left(\frac{T^i_{m^i} - T^i_{m^i - 1}}{\alpha^i} \cdot 2\pi\right) > 0 \quad , \quad 0 < J^i_{m^i} < \frac{\alpha^i}{2} \tag{E.7}$$

So we need to solve for α^i that maximizes (E.7), which may be rephrased as trying to get $(T^i_{m^i} - T^i_{m^i-1})/\alpha^i$ to be as close to an integer as possible.

NHM is based on this observation and attempts to find α^i in a gradient-descent manner, by searching for the set of best fitting integers to normalise the distances between elements in T^i_{RX}. The step-wise procedure of NHM is described by Alg. 1.

Algorithm 1: ©2020 IEEE: Normalised Harmonics Mean (NHM)

Data: dT^i_{RX}
Result: α^{i*}
1 Make a preliminary hypothesis for α^{i*}.
2 **while** α^{i*} *has not converged* **do**
3 Estimate $\eta^i_{m^i}$ for α^{i*}.
4 Update the hypothesis for α^{i*} based on $\eta^i_{m^i}$.
5 Calculate KPIs based on app parameters.

Let the set dT^i_{RX} denote the set of the distances between neighboring elements T^i_{RX}. The preliminary hypothesis for α^i could be $\alpha^{i*} = \mathrm{mean}(dT^i_{\mathrm{RX}})$.

Outage creates harmonic contributions of orders higher than 1, $o^i_{m^i} > 0$, in the set dT^i_{RX} as depicted in Fig. E.3. In the next step we seek to normalize these harmonics by estimating the series of latent variables $\eta^i_{m^i}$ given by (E.8). We do this under the condition that $||J|| < \dfrac{\alpha^i}{4}$ (or equivalently $0 \leq J < \dfrac{\alpha^i}{2}$) such that aliasing is avoided, $n^i_{m^i} \leq n^i_{m^i+1} + 1$, which in practice means that a transmission is never received earlier than a previous transmission.

$$\eta^i_{m^i} = \left\lfloor \frac{T^i_{m^i} - T^i_{m^i-1}}{\alpha^{i*}} \right\rceil \tag{E.8}$$

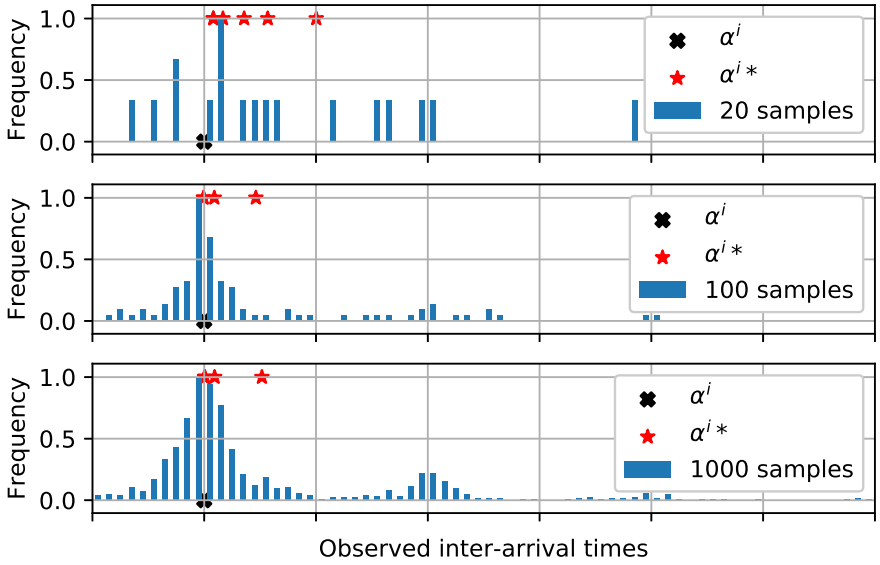


Fig. E.3: ©2020 IEEE: Histogram of dT^i_{RX} where $J^i_m \sim \exp(.25)\cdot\alpha^i/2$ and $p_{\mathrm{o}} = 0.2$. The red crosses indicate the iterations of α^{i*} in NHM, which can be seen to approach α^i, even when few samples are available. In this example, the initial hypothesis of α^{i*} was based on the mean of dT^i_{RX}.

where $n^i_{m^i} = \sum_{k=1}^{m^i} \eta^i_k \quad \Leftrightarrow \quad \eta^i_m = n^i_m - n^i_{m-1}$.

The initial hypothesis for α^{i*} is unlikely to be correct, but it serves to provide an initial estimate of the latent series $\eta^i_{m^i}$, which we can now use to estimate α^i. Given the hypothesis that α^i is the mean of a normalised version of dT^i_{RX} we have (E.9).

$$\alpha^{i*} = \frac{1}{M^i} \sum_{m^i=2}^{M^i} \left(\frac{T^i_{m^i} - T^i_{m^i-1}}{\eta^i_{m^i}} \right) \tag{E.9}$$

Iterative updates of $\eta^i_{m^i}$ and α^{i*} Eq. (E.9) will gradually go towards α^i as depicted in Fig. E.3. The required number of iterations for convergence depends on the accuracy of the initial estimate of α^i, which is dependent on p_{o} and the number of samples. After the algorithm convergences the estimated number of transmissions lost between successive receptions is given by the latent variables $o^i_{m^i} = n^i_{m^i} - n^i_{m^i-1}$. Furthermore β^i can be found by linear regression of (E.2) as both α^i and $n^i_{m^i}$ have been estimated, but is not necessary for computing the KPIs in our case.

NHM can be run in an online manner by saving dT^i_{RX} and T^i_{M+1} and updating both upon reception of a new transmission for app i before repeating step 2 starting from the previous α^{i*}.

3.2 Classification

In case the parameters α^i and β^i are known for all applications $i \in I$, but transmissions are not labeled by their parent application, we need to classify which application a new reception belongs to. Let $f(T^i_{\mathrm{RX}}, \alpha^i, \beta^i)$ be the transmission generating function, then we may propose a periodic likelihood function $g(T^i_{\mathrm{RX}}, \alpha^i, \beta^i)$, such that

$$f(T^i_{\mathrm{RX}}, \alpha^i, \beta^i) \propto g(T^i_{\mathrm{RX}}, \alpha^i, \beta^i) = \frac{\cos\left(\frac{T^i_{\mathrm{RX}} - \beta^i}{\alpha^i} \cdot 2\pi\right) + 1}{2} \tag{E.10}$$

Then i is found by $\max_{i \in \{1:I_{\mathrm{apps}}\}}(g(T^i_{\mathrm{RX}}, \alpha^i, \beta^i))$. Here, we chose a cosine function as a likelihood function and added 1 before normalizing to keep the probability between 0 and 1. In practice the maximization yields the same result regardless of these linear operations due to the associative property of the periodic likelihood function.

A sub-problem here is the initial classification when β^i is not known, while α^i is known, we must find the most likely sequence of packets to belong to the process that has the inter-arrival time α^i for some β^i. Then $g(T^i_a - T^i_b, \alpha^i, \beta^i = 0)$ is the likelihood that transmissions a and b belong to the same sequence with inter-arrival time α^i. If we then let transmission m be a newly received packet and m^i be the last packet received by app i, then we can simply find i by $\max_{i \in \{1:I_{\mathrm{apps}}\}}(g(T^i_m - T^i_{m^i}, \alpha^i, \beta^i = 0))$. Notice, that errors in the estimate of α will results in a error in the output of g, which increases as the distance between two timestamps increases, or in other words, labeling is more likely to be erroneous as p_{o} increases or after offline periods.

4 Clustering

When traffic parameters are unknown and the traffic is unlabeled by its parent application we need to perform clustering of the received traffic. In this section we propose a clustering algorithm that estimates clusters in a hierarchical manner based on the Lomb-Scargle periodogram and an online version, which assigns newly received transmissions to already known clusters in a greedy manner.

Let the set of all unassigned traffic received by device x be T^{x*}_{RX}.The procedure for the hierarchical clustering method, Successive Periodicity Clustering (SPC) is described in Alg. 2.

NHM can be used for creating a hypothesis for α^i. This is a robust approach as long as $\alpha^1 << \alpha^2 << ... << \alpha^{I_{\mathrm{apps}}}$ with an estimation error that increases greatly as $\frac{\alpha^i}{\alpha^{i+1}} \to 1$. However, it may often be the case that $\frac{\alpha^i}{\alpha^{i+1}} \approx 1$ so we will examine using the Lomb-Scargle periodogram for hypothesis creation instead.

Lomb in [13] proposed a least-squares periodogram, which tested for the best fitting set of frequencies for a unevenly sampled time-series. In [14], Scargle proposed a generalised

Algorithm 2: ©2020 IEEE: Successive Periodicity Clustering (SPC)

Data: T^{x}_{RX}
Result: $\{T^{i*}_{\mathrm{RX}}, \alpha^{i*}, \beta^{i*}\}, T^{i}_{\mathrm{RX}} \in T^{x}_{\mathrm{RX}}$
while do
  Hypothesize α^{i*} for an app i, given T^{x*}_{RX}.
  if α^{i*} *is significant* **then**
    Label data points that best fit α^{i*} and extract β^{i*} and T^{i*}_{RX}.
    Calculate KPIs based on α^{i*}, β^{i*} and T^{i*}_{RX}.
    Remove T^{i*}_{RX} from the set of unlabeled data. $T^{x*}_{\mathrm{RX}} = T^{x*}_{\mathrm{RX}} \backslash T^{i*}_{\mathrm{RX}}$
  else
    return $\{T^{i*}_{\mathrm{RX}}, \alpha^{i*}, \beta^{i*}\}, T^{i}_{\mathrm{RX}} \in T^{x}_{\mathrm{RX}}$

form of the classical Fourrier periodogram, which turns out to be equivalent to the least-squares fitting of sinusoids from [13]. Hence, this periodogram was termed *Lomb-Scargle periodogram* and [15] treats its statistical properties, including a close-upper limit on the probability of falsely detecting a peak frequency in a data set comprised solely of noise. The Lomb-Scargle periodogram is well-known in some scientific communities, but was recently introduced to a wider audience in [16] by surveying works on and related to the Lomb-Scargle periodogram and lending conceptual intuitions. The Lomb-Scargle algorithm has been implemented in Python in the Astropy package [20, 21].

4.1 Lomb-Scargle-based hypothesis creation for α

The classical Fourier periodogram requires evenly sampled data, but the Lomb-Scargle algorithm can find a PSD-like density for unevenly spaced times series [13, 14, 16].

Preprocessing of our data is required. We generate a series V which is the same size as T^{x*}_{RX} and holds the value 1 for each of the timestamps. Then we interpolate V with 0's to permit sine-based analysis. We carry out this interpolation in a heuristic manner by finding places in dT^{x*}_{RX} larger than dT^{x*}_{RX} in an attempt to avoid interpolation between transmissions that are very close in time, i.e., to avoid giving credibility to very high frequencies.

We must identify the frequency spectrum that is relevant for analysis. The size of the set of frequencies within the grid dictates both accuracy and computational effort, since 'false positive' local peaks in the periodogram are less likely to be found to be the global peak, but at the expense of evaluating the periodogram in more points. The minimum detectable frequency will be a signal that completes one oscillation over the entire period of the data-set, $f_{\min} = \frac{1}{T^{x*}_{M^{x*}} - T^{x*}_{1}}$. The spacing of the frequency grid, $\delta f = \frac{1}{n_0 \tau_{\mathrm{w}}}$, depends on an oversampling factor, n_0. The higher n_0 is, the higher the chance that a

peak frequency is not missed in the analysis. Typically, 8 is used [16]. The Nyquist limit does not always exist for the unevenly sampled Lomb-Scargle periodogram and in our case it will inevitably be very large since $f_{Ny} = 1/(2p)$, where p is the largest value that all T^{x*}_{RX} can be written as an integer multiple of [22]. Instead we shall use a heuristic to determine $f_{\max}$. Say that I_{apps} applications are generating packets with rates α^i and there is no outage. In this case the mean observed inter-arrival time is always less than or equal to $\frac{1}{M^i-1}\sum_{m^i=1}^{M^i-1}(dT^i_{m^i}) \leq \min_{i\in\{1:I_{\mathrm{apps}}\}}(\alpha^i)$. Then if we choose $f_{\max}$ such that

$$f_{\max} = \frac{L}{\frac{1}{M^{x*}-1}\sum_{m^{x*}=1}^{M^{x*}-1}(dT^{x*}_{m^{x*}})} \tag{E.11}$$

where L is scalar used to take into account the effect of outage. Given 50% outage in a set of times one would expect to find a mean inter-arrival time that is twice as long. So assuming L = 2, our heuristic limit for the frequency grid should cover the maximum frequency of any observed app for $p_{\mathrm{o}} \leq 50\%$.

Let $Z = P(f)$ be the periodogram value at frequency f. In [14] it was observed that the probability of observing a periodogram value less than Z in pure Gaussian noise can be expressed as (E.12).

$$P_{\mathrm{single}}(Z) = 1 - \exp(-Z) \tag{E.12}$$

A close-upper limit for the false alarm probability, FAP(Z), was found by Baluev in [15]. We access this limit at the peak-value of the periodogram, Z_0, with significance, σ and only accept the hypothesis $\alpha^* = 1/f_0$ if FAP(Z) $< \sigma$.

4.2 Labeling and collision resolution

If a significant peak is detected we need to label data that fits the suggested α^{i*}. Since up to I_{apps} apps have generated T^x_{RX} we must attempt to find the sequence of data points that is the best fit for our proposed application i. We do this by labeling transmissions, which fit coarsely and then sorting out incoherent false positive transmissions. Given a significant proposal for α^{i*} the procedure is:

1. Mark transmissions that seem to be a good fit for α^{i*}.
2. Remove marks from transmissions that are incoherent with the rest of the labeled transmissions.
3. Label marked transmissions as belonging to app i.
4. Calculate KPIs based on app parameters and labeled packets.

We calculate the fit between all times in all possible sequences by building a matrix $\mathbf{IA} = \mathbf{T^{x*}_{RX}U} - \mathbf{T^{x*}_{RX}U}^T$ where $\mathbf{U}$ is a vector of ones [1,1,1,...,1] of length M and $\mathbf{T^{x*}_{RX}}$ is a vectorized version of the set of unlabeled transmissions. We then proceed to estimate the fit, ϕ_m, of each transmission to all other transmissions given α^{i*} as (E.13).

$$\phi_m = \sum_{n=1}^{M} \left(g(\mathbf{IA}_{(m,n)}, \alpha^{i*}, \beta^i = 0) \right) \tag{E.13}$$

Now, a preliminary sorting of transmissions is performed by labeling only transmissions for which $\phi_m > \frac{1}{2M^{x*}} \sum_{m=1}^{M^{x*}} (\phi_m)$. This coarse sorting of packets results in a fair amount of false positives in T^{i*}_{RX}. False positives can in many cases (especially when p_{o} is low) be identified as clearly incoherent transmissions that 'collide' in time with other transmissions in the set. The minimal period between packets is $\frac{\alpha^{i*}}{2}$ before a pair of transmissions are considered to collide. This is conditioned by $||J^i|| < \frac{\alpha^{i*}}{4}$. Colliding pairs are resolved by checking the fit of the colliding packets and removing the worst fitting transmission from the set of transmissions for the new app. Given a collision at T^x_l and T^x_{l+1} we have that

$$C_{l \in \{1:M^i\} | app^i} : \phi_l > \phi_{l+1} \tag{E.14}$$

Now we have estimated T^i_{RX} from the set of unlabeled data. Since the Lomb-Scargle method is only as accurate as the frequency grid we use in its analysis, we can attempt to use NHM on the estimated set T^i_{RX}, to get better estimate for α^i. Then we can compute KPIs for the application and finally remove the labeled traffic from T^{x*}_{RX} and attempt SPC again.

4.3 Greedy online clustering

SPC does batch processing and is relatively computationally heavy. An online version of the algorithm that minimizes the computational effort and can be used for real-time monitoring is warranted. So we introduce a greedy online clustering (GOC) algorithm. Upon the reception of a new transmission $T^x_{M^X}$ GOC runs as described in Alg. 3.

Once a new transmission is received, it should be checked if the reception time fits with the expected reception time of any known applications. GOC estimates $\phi^i = g(T^x_{M^X} - T^i_{M^i}, \alpha^{i*}, \beta^i = 0)$ for all known i. Then assigns the transmission to T^i_{RX} if $\max_{i \in \{1:I_{\mathrm{apps}}\}}(\phi^i)$ is positive. Then the KPIs are updated for app i.

In case $\max_{i \in \{1:I_{\mathrm{apps}}\}}(\phi^i) \leq 0$ the new transmission is added to the set of unlabeled transmissions T^{x*}_{RX} and SPC is attempted.

Algorithm 3: ©2020 IEEE: Greedy Online Clustering (GOC)

Data: T^x_{MX},$\{T^{i*}_{\mathrm{RX}}, \alpha^{i*}, \beta^{i*}\}$, $T^i_{\mathrm{RX}} \in T^x_{\mathrm{RX}}$
Result: $\{T^{i*}_{\mathrm{RX}}, \alpha^{i*}, \beta^{i*}\}$, $T^i_{\mathrm{RX}} \in T^x_{\mathrm{RX}}$
Estimate the fit ϕ^{i*} of the new arrival T^x_{MX} for every previously known app i.
if $\max(\phi^i) > 0$ **then**
    Add the transmission T^x_{MX} to the app j which had the best fit.
    Update the parameters, α^{i*} , β^{i*}.and the KPIs for i.
else
    Add the transmission to the set of unlabeled transmissions T^{x*}_{RX}. and run SPC on T^{x*}_{RX}.

5 Probabilistic Performance

In this section numerical results for the performance of the algorithms introduced throughout this paper are presented. The performance is evaluated in terms of the accuracy of the outage estimation and offline state detection.

All the presented algorithms and the naïve method presented in Sec. 2.4 have been implemented in Python. The implementations are based on functionality from the Numpy package [23] and the Lomb-Scargle algorithm of the Astropy package [20, 21]. A parameter ψ that reduces the computational requirements and enhances the accuracy of SPC and GOC has been introduced. ψ ensures that SPC is only ever run when T^{x*}_{RX} is larger than ψ. Furthermore after labeling and collision resolution the size of the proposed set T^i_{RX} is checked and only accepted if it is larger than ψ.

The performance has been evaluated against network outage probabilities, $p_{\mathrm{o}} = [0, .1, .2, .3, .5]$. For each data point 1000 stochastic transmission sequences with jitter and outage have been generated. Outage has been randomly induced in the sequences by generating samples for T^i_{RX} and discarding $T^i_{\mathrm{m}^i}$ with probability p_{o} until the specified sample size was reached. Sample sizes of 5, 10, 25, 50 and 100 transmissions have been used. The traffic parameters are drawn from the following distributions:

$$\alpha^1 \sim U(100, 200)$$
$$\alpha^2 \sim U(1, 5) \cdot \alpha^1$$
$$\beta^i \sim U(0, 0.5) \cdot \alpha^i$$
$$j^i_m \sim \exp(0.2) \cdot \frac{\alpha^i}{20}$$

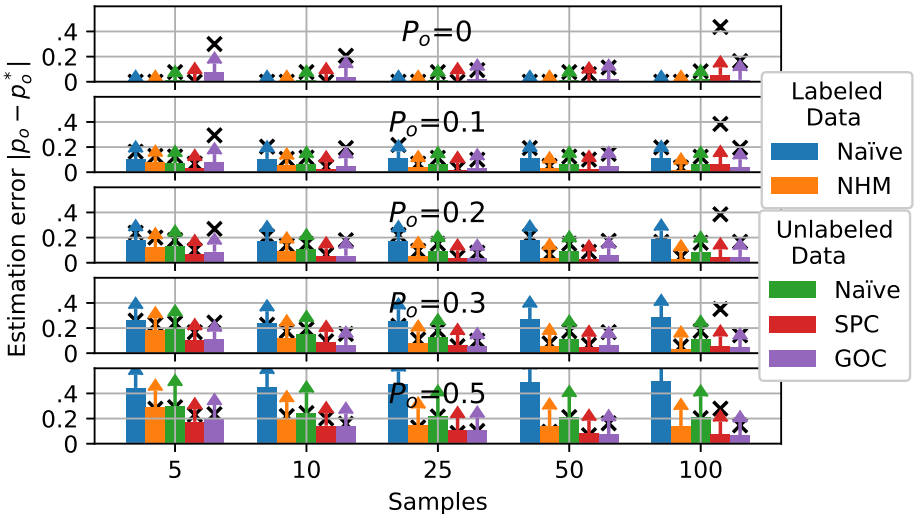


Fig. E.4: ©2020 IEEE: Absolute error in network outage estimation for NHM, GOC, SPC and the naïve approach. The naïve method is applied to both the labeled data-set and the unlabeled data-set ($\tau_w = 1500$). The bars indicate the mean absolute errors, the arrows indicate the size of the standard deviation and the 95th percentile is denoted by the black crosses.

5.1 Network outage detection and estimation

The absolute estimation error, $|p_o - p_o^*|$, is plotted in Fig. E.4 for NHM and the naïve method on a labelled sequence and SPC, GOC and the naïve method for unlabelled sequences. The labelled sequence is T^1_{RX}, and the unlabelled sequence is $T^x_{\mathrm{RX}} = T^1_{\mathrm{RX}} \cup T^2_{\mathrm{RX}}$ and $\psi = \min(s_{\min}, 25)$ for SPC and GOC, where $s_{\min} = \min(M^1, M^2)$.

The naïve approach works well when there is little to no uncertainty, but both the mean error and the variation in error increases rapidly as p_o increases. Notice that NHM has both a lower mean and less variance in the estimation error, especially as the number of available samples increase. Indeed, NHM has a <5% estimation error at $p_o = 0.3$ for 50 available samples. On unlabeled data-sets, The naïve approach is found to perform a little better, which is expected due to the increased number of samples within the 'windowing' function of the naïve approach. Still, both SPC and GOC outperforms the naïve approach as p_o increases. We observe that SPC performs a little worse at 100 samples compared to 50 samples - this can be explained by the periodic labelling function performing worse over such a long period of samples, which is exaggerated when SPC then tries to find an 'imaginary' application to label transmissions that should have been assigned previously. In GOC, the inaccuracy of the periodic labelling function causes errors when only a few samples are available, here 'false positive' labels result in the secondary application not finding sufficient samples to satisfy ψ.

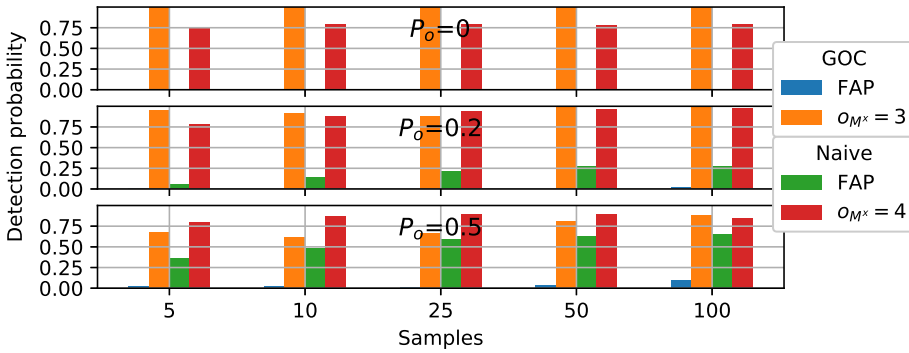


Fig. E.5: ©2020 IEEE: Probabilities of detecting offline-state in different conditions for the real-time clustering method, GOC, and the naïve approach ($\epsilon = .25$, $\tau_w = 1500$). FAP denotes the false alarm probability. o_{M^x} denotes the number of consecutively missed transmissions during the offline period.

5.2 Offline state detection

In this scenario the device goes offline and we denote the number of transmissions since the device went offline, o_M^x. We define the false alarm probability (FAP) as the probability that an offline state is detected when the device is not offline. In Fig. E.5 the resulting probabilities are plotted. We observe that GOC identifies devices as offline correctly with a high probability. The probability of correctly identifying offline devices increases with the number of sampled reception times. The FAP for GOC increases with the network outage, as α^{i*} is more prone to be inaccurate, which is also the case when a very small number of samples are available before the device goes offline. Furthermore, when the network outage is high, the device is more likely to appear offline due to random outage in the received sequence T_{RX}^x, which affects the naïve method significantly more than GOC.

5.3 Sampling and Computation time

The total computation time of each algorithm is plotted in Fig. E.6. The computation time of GOC is significantly larger than that of SPC, since GOC will attempt SPC multiple times as transmissions are received, however, one should note that the cmputing time of GOC when receiving transmissions from known apps is very low, near that of NHM. The computation time of NHM is significantly lower than that of any other method. The naïve approach beats both clustering methods in terms of computation time. Still, in the special case that $\min(\alpha^i)$ is 1 hour for a group of devices, GOC supports initial analysis of up to 12240 devices in serial on a single thread of a 4.00 GHz i7-6700 if 100 samples are received as batch for each device, but processed in an online manner using GOC (worst-case). In practice, the sampling time for MAR-P traffic [18]

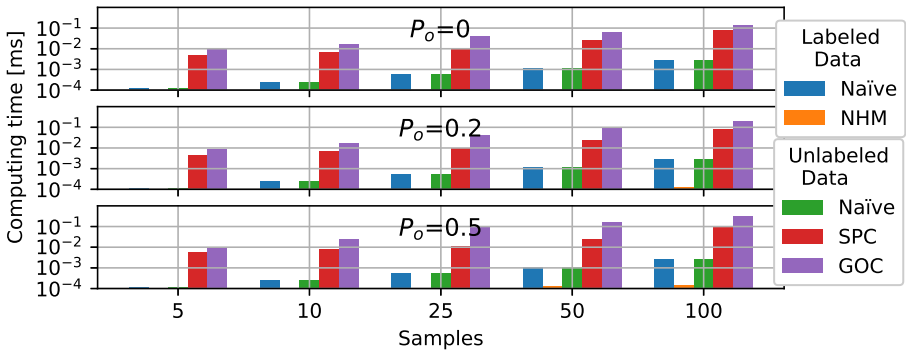


Fig. E.6: ©2020 IEEE: Mean computation time for NHM, SPC, GOC and the naïve approach. Mind that the y-axis is logarithmic.

would be much greater, such that even more devices could be supported.

6 Case study: smart metering deployment

This section presents monitoring results from using GOC on a data-set from a real-world wireless smart metering deployment. The data consists of timestamps, device IDs and IDs of the receiving concentrators for a total of 1048576 received transmissions from 4522 smart meters. The smart meters are deployed in a large area with 6 concentrators that provide coverage for all meters. Meters are connected to concentrators in a star-topology.

Meters broadcast quasi-periodically to the concentrators at an inter-arrival time that is normally distributed around three hours. It should be noted that meters, which do not receive an acknowledgement will include the data of any missed transmissions in their broadcast. Effectively this means that the app layer QoS is kept high even in poor link level conditions. Here, we monitor the network performance at the link level from the perspective of individual concentrators and the perspective of a centralised service gathering data from all concentrators.

The distribution of estimated α^{i*}'s using GOC and NHM are plotted in Fig. E.7 for the perspective of a central server and as a joint distribution for all concentrators. Here we find that the estimated $\alpha^{*}i$'s are normally distributed around 3 hours for GOC, which is coherent with the deployment case. NHM on the other hand shows a fat tail in the distribution, which is not coherent with our prior knowledge of the transmission rate distribution. This inconsistency can be ascribed to the number of available data, noise in the data and the inaccuracy of NHM at increasingly higher network outages as seen in Fig. E.4.

Overlapping concentrator coverage enhances the application level-performance; Using

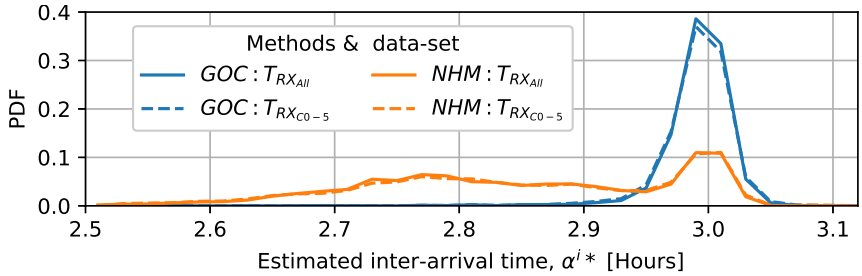


Fig. E.7: ©2020 IEEE: Distributions of estimated inter-arrival times from the perspectives of each concentrator and a centralised server.

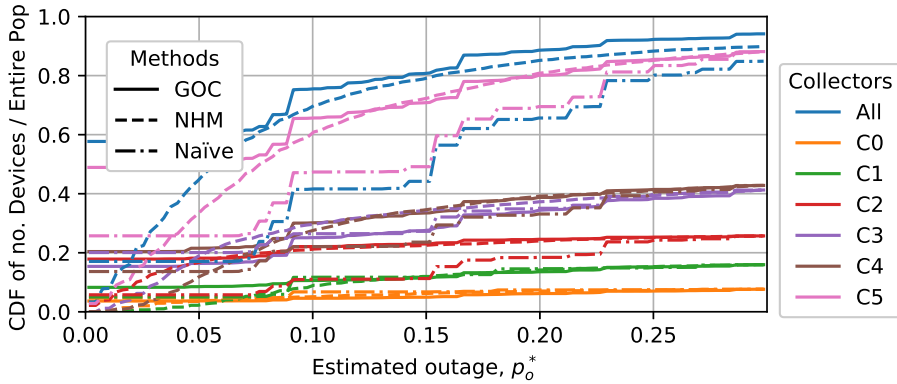


Fig. E.8: ©2020 IEEE: CDFs of estimated outages for each device as seen by each concentrator and a centralised server. The CDFs are evaluated against the entire population of devices in the deployment.

GOC, we find that 20.3% of the meters are connected to at-least one concentrator without outage at the link-level, while the same measure is 54.8% from the perspective of a centralised server. The PDF of estimated outages for all meters from the perspective of the concentrators and the centralised server can be found in Fig. E.8. Each given device is not necessarily in the range of each deployed concentrator, such that, as expected, we see varying connectivity levels for the different concentrators. The estimated outage at the link level as seen by the server is much lower than that of any individual concentrator. We have used a parameter of $\psi = 10$ for the analysis. This means that 3.4% of the devices represented in the data were not analysed due to being sampled less than 10 times while another 1.1% did not exhibit clear periodicity due to having few samples and likely a relatively high outage. This group of devices are clear outliers in the deployment and warrant further investigation.

The distribution of the estimated outage by NHM and GOC converges after $p_0^* = 5\%$.

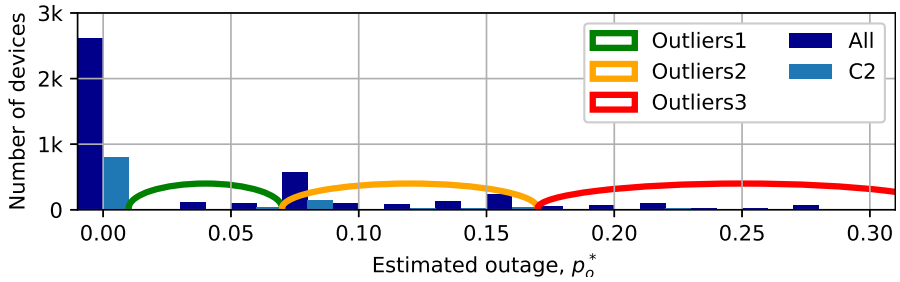


Fig. E.9: ©2020 IEEE: Histogram of estimated outages for devices in the deployment at the central server and concentrator C2. Devices can be grouped in spectra, which increase in width in proportion to the estimated outage, or equivalently, as reliability of the estimate goes down.

Both methods identify roughly the same sets of devices for different groups of outliers. The naïve method continually overshoots it's outage estimation. In Fig. E.9 we take a closer look at the links experienced by the concentrator "C2" and the links as experienced by the server. Here a histogram of the estimated outage at the link-level is depicted. In this way groups of outliers can be identified both at the application level and at the level of individual concentrators. In summary, the methods can quantify the coverage of concentrators in practice, which enables evaluation of their placement. The locations of concentrators and devices would be known by the smart metering company, making it relatively easy to assess spatial correlation of poorly performing devices. Installing a local concentrator is sensible in areas with many poorly performing devices, whereas antenna upgrades might be more cost beneficial for solitary, poorly performing devices. The methods can also help identifying Byzantine transmissions and seasonality in the outage.

7 Conclusion

In this paper we introduced methods for passive detection in IoT networks in deployments with quasi-periodic reporting, such as smart-metering, environmental monitoring and agricultural monitoring. The methods are applicable in both mesh networks and LPWA and cellular networks, setting them apart from the state-of-art methods for fault detection in WSNs.

The SPC and GOC algorithms were shown to perform well even at high outage for composite sequences, which makes them well suited for monitoring devices and networks in 'black-box' networks where the outage may be quite high. The cost of the utility and precision of these methods is computational effort, which was measured for a 4.0GHz i7-6700 CPU. Furthermore, the utility of NHM and GOC has been exhibited through a short analysis of real-world data from a smart meter deployment.

Acknowledgments

This work was in part supported by a grant through FORCE Technology from the Danish Agency for Institutions and Educational Grants, in part supported by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (ERC Consolidator Grant Nr. 648382 WILLOW) and Danish Council for Independent Research (Grant Nr.8022-00284B SEMIOTIC).

Symbol	Meaning	Symbol	Meaning
I^x_{apps}	The number of applications running on device x.	m^i	The index of the received packets from application i.
τ_{w}	A time window.	T^x_{TX}	The set of all transmissions from device x.
T^i_m	The reception time of the m'th packet from application i.	$\eta^i_{m^i}$	The harmonic order of an observed inter-arrival rate. Used in NHM.
T^i_{TX}	The set of all transmissions from application i on device x.	α^i	The period between transmission generation for quasi-peridic application i.
*	The star denotes estimated values.	T^x_{RX}	The set of all received transmassions from device x.
β^i	The offset in the period of application i.	$f_{\min}, \delta f$	The corner- and step-frequencies used in Lomb-Scargle.
dT^x_{RX}	The observed inter-arrival period for device x.	$J^i_{m^i}$	The E2E delay between generating a transmission to reception.
IA	A matrix of all potential inter-arrival times for a set, T^x_{RX}.	T^i_{RX}	The set of all received transmissions from application i.
C_{off}	A classification of being offline.	ϕ_m	The likelihood of transmission m being a part of a given application.
dT^i_{RX}	The observed inter-arrival period for application i.	p_{o}	The network outage
Ψ	The minimal number of observations in an application.	$o^i_{m^i}$	The cumulative number of missed receptions up to index m from application i.
k	The limit of consecutively missed transmissions before a device is classified as being offline.	$FAP(Z)$	A likelihood estimator for packet m to belong to a specific application.
n^i	The index of the transmissions from application i.		

Table E.2: ©2020 IEEE: Index of notations and variables

References

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, Fourthquarter 2015.

[2] S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "Nr: The new 5g radio access technology," *IEEE Communications Standards Magazine*, vol. 1, no. 4, pp. 24–30, Dec 2017.

[3] K. Mekki, E. Bajic, F. Chaxel, and F. Meyer, "A comparative study of lpwan technologies for large-scale iot deployment," *ICT Express*, vol. 5, no. 1, pp. 1–7, 2019.

[4] S. Popli, R. K. Jha, and S. Jain, "A survey on energy efficient narrowband internet of things (nbiot): Architecture, application and challenges," *IEEE Access*, vol. 7, pp. 16 739–16 776, 2019.

[5] R. Ratasuk, N. Mangalvedhe, D. Bhatoolaul, and A. Ghosh, "Lte-m evolution towards 5g massive mtc," in *2017 IEEE Globecom Workshops (GC Wkshps)*, Dec 2017, pp. 1–6.

[6] A. Mahapatro and P. M. Khilar, "Fault diagnosis in wireless sensor networks: A survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 2000–2026, Fourth 2013.

[7] Z. Zhang, A. Mehmood, L. Shu, Z. Huo, Y. Zhang, and M. Mukherjee, "A survey on fault diagnosis in wireless sensor networks," *IEEE Access*, vol. 6, pp. 11 349–11 364, 2018.

[8] M. Nazari Cheraghlou, A. Khadem-Zadeh, and M. a. Haghparast, "A framework for optimal fault tolerance protocol selection using fuzzy logic oniot sensor layer," *International Journal of Information & Communication Technology Research*, vol. 10, no. 2, 2018. [Online]. Available: http://ijict.itrc.ac.ir/article-1-326-en.html

[9] Y. Liu, K. Liu, and M. Li, "Passive diagnosis for wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 4, pp. 1132–1144, Aug 2010.

[10] B. C. Lau, E. W. Ma, and T. W. Chow, "Probabilistic fault detector for wireless sensor network," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3703 – 3711, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0957417413009548

[11] X. Jin, T. W. S. Chow, Y. Sun, J. Shan, and B. C. P. Lau, "Kuiper test and autoregressive model-based approach for wireless sensor network fault diagnosis," *Wireless Networks*, vol. 21, no. 3, pp. 829–839, Apr 2015. [Online]. Available: https://doi.org/10.1007/s11276-014-0820-0

[12] R. F. Stellingwerf, "Period determination using phase dispersion minimization." *Astrophysical Journal*, vol. 224, pp. 953–960, Sep 1978.

[13] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophysics and Space Science*, vol. 39, no. 2, pp. 447–462, Feb 1976.

[14] J. Scargle, "Studies in astronomical time series analysis. ii - statistical aspects of spectral analysis of unevenly spaced data," *The Astrophysical Journal*, vol. 263, 01 1983.

[15] R. V. Baluev, "Assessing the statistical significance of periodogram peaks," *Monthly Notices of the Royal Astronomical Society*, vol. 385, no. 3, pp. 1279–1285, April 2008.

[16] J. T. VanderPlas, "Understanding the lomb–scargle periodogram," *The Astrophysical Journal Supplement Series*, vol. 236, no. 1, p. 16, may 2018.

[17] O. Al-Khatib, W. Hardjawana, and B. Vucetic, "Traffic modeling and optimization in public and private wireless access networks for smart grids," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1949–1960, July 2014.

[18] 3GPPP, "Cellular system support for ultra-low complexity and low throughput internet of things (ciot)," 3GPP, Technical report (TR) 45.820, 12 2015, version 13.1.0.

[19] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb 2014.

[20] Astropy Collaboration and T. P. e. a. Robitaille, "Astropy: A community Python package for astronomy," *Astronomy and Astrophysics*, vol. 558, p. A33, Oct. 2013.

[21] Astropy Contributers and A. M. e. a. Price-Whelan, "The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package," *Astronomical Journal*, vol. 156, p. 123, Sep. 2018.

[22] L. Eyer and P. Bartholdi, "Variable stars: Which Nyquist frequency?" *Astronomy and Astrophysics, Supplement*, vol. 135, pp. 1–3, Feb. 1999.

[23] T. E. Oliphant, *A guide to NumPy*. Trelgol Publishing USA, 2006, vol. 1.

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-817-9

AALBORG UNIVERSITY PRESS